

Berlin and global warming policy

The first serious test of the Rio Convention on Climate Change will come later this month, when the members of the treaty decide what happens next. They need above all to stay cool.

REMEMBER the treaty signed at Rio de Janeiro in 1992 to contain the threat of global warming caused by greenhouse gases in the sky? For a few months, the treaty gave the world a sense that something had been done. But a piece of paper does not, by itself, reduce atmospheric carbon dioxide concentrations. The real business of the treaty will begin only this month, with the Conference of the Parties due to open at Berlin on 28 March. It will probably be a rowdy affair. Following now standard routine at UN conferences, there will be a full representation of what are called non-governmental organizations (NGOs), many of which will make a great deal of noise, and grab a slew of headlines, with predictions of imminent catastrophe and complaints that governments are complacent. Governments themselves will differ on the reality of the threat, its immediacy and the best means of avoiding it. What will, or should, emerge?

First, there is or should be no doubt that the threat of global warming is a reality. That assertion does not rest on measurements of temperature over recent years, on reports that the height of the sea-surface is increasing by about 1 mm a year or even, in the past few weeks, on reports that ice shelves on the Antarctic peninsula are breaking up. Rather, it rests on what is known and understood about the Earth's atmosphere and its function as regulator of temperature on the surface. It is common ground that if there were no greenhouse gases in the atmosphere, the average temperature on the surface would be at freezing or below. Ice-core measurements show that the carbon dioxide concentration has increased by 25 per cent since 1800, and that almost a half of that increase has been concentrated in the past three decades. But if 280 parts per million by volume (p.p.m.) made the Earth warm enough to be habitable in 1800, is it not reasonable to expect that 355 p.p.m. now will have further increased the average temperature? And that the further increases in prospect will do more?

Trends

That is why the threat of global warming must be taken seriously. That case is strengthened by the emergence in this century of other artificial greenhouse gases (methane, nitrous oxide and the chlorofluorocarbons in particular) which are, in principle, significant modulators of surface temperature, and at present account for roughly 40 per cent of the greenhouse effect. It is true, of course, that these trends do not prove that the temperature at the surface will continue to increase more or less in step with the degree of

climate forcing attributable to greenhouse gases. Where climate research is now stuck is in the attempt to model the responses more accurately.

Models

Attempts so far to model the effects of the greenhouse gases are imperfect, to say the best of them. The continuing neglect of atmospheric water, both as vapour (which is a greenhouse gas) and as ice (which may, in clouds, increase the Earth's albedo), is an obvious weakness. Given the complexity of the atmosphere, an altered heat balance in the atmosphere may disproportionately enhance radiation from the upper atmosphere, creating the built-in air-conditioner on which some have speculated. Or there may be other Gaia-like mechanisms that will save the day. But the onus of proof has now shifted from those who advocate global warming as a threat to those who hold that it may not be.

But how great and how urgent is the threat? That is what governments demand to know. Sadly, there are no simple answers. An increase of the average surface temperature of one degree, comparable with the kinds of change affecting particular parts of the Earth's surface in recorded history, would transform life on the surface of the Earth, but in ways that cannot be forecast accurately; modelling regional climate change is necessarily a refinement of global modelling. The timescale of changes of that order is similarly incalculable, but half a century might see a one-degree increase of temperature. That is not long, given the glacial pace of international diplomacy.

The NGOs fighting for an audience at Berlin will have more scary tales to tell. What if the models have underestimated the response of the Earth's surface? What if the 50 per cent of the carbon dioxide released each year is buried in the oceans, and is suddenly released? Such evidence as there is suggests burial in humus and as carbonates. What if the West Antarctic ice sheet should slump into the Southern Oceans? The question points to the need for monitoring, not at this stage for panic. But none of this argues for complacency. The scale of the threat of global warming, and its timing, will become plain only by careful monitoring of present trends on the surface of the Earth. That is one issue that cries out for attention at Berlin: how will the monitoring be done, and who will interpret the data?

For the research community, the last is the crucial question. The running so far has been made by the Intergovernmental Panel on Climate Change (IPCC), created some



years ago by the World Meteorological Organization (WMO) and the UN Environment Programme (UNEP). In the best part of a decade, it has engaged hundreds of able people in its work. It will produce a third assessment of the climate situation at Berlin, where it apparently hopes to be accepted as the Rio treaty's technical instrument in perpetuity. That is not a good idea. IPCC, for all its many virtues, also reflects the weaknesses and vested interests of its sponsors. And, as on a famous occasion last year, it can easily become the prisoner of the agencies' publicity machines.

Assessment

This journal has argued almost from the outset that the United Nations has a better model to follow, that of the UN Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), which has over several years produced a series of technical assessments that carry general conviction. The assessment part of the climate job requires, at the outset, a steering committee, a full-time scientific staff (preferably on secondment), a small army of paid consultants and a mechanism whereby components of its assessments can be discussed by small seminars of knowledgeable people. Its primary aim should be to review the whole field of climate change at regular intervals, say once a year. It should be content with assessments written in technical language; governments have plenty of technical people to advise them of the importance of what emerges.

The same organization should also be competent to take account of the socio-economic trends likely to affect both the input to global warming (fuel usage) and the output (migration, for example). That is what the United Nations needs. The notion that such an organization could be based at the International Institute for Applied Systems Analysis at Laxenburg, near Vienna, is not absurd. There is also a need for an organization that will systematically update a list of the questions to which answers are needed urgently. The risk, in the monitoring of climate trends, is that people will look at the same phenomena everywhere because they offer the least difficulty.

Criticism

IPCC will declare this implied criticism to be unfair. It publicly identifies the chairmen of its several topic panels, with whom dissenters are invited to communicate, and sends out drafts of its documents for 'peer review' before they see the light of day. But that does not guarantee that its materials are fully comprehensive and welcoming of contrary opinions. They reflect the received wisdom in the nicest way, as if an argument that the Earth's atmosphere is its own air-conditioner is best dealt with by being ignored. It would be a different story if IPCC's two assessments so far had listed the various (mostly crackpot) suggestions there have been so far for ridding the atmosphere of greenhouse gases.

The creation of an efficient monitoring and assessment procedure is one step that the treaty members should take at their meeting in Berlin later in the month. But the real arguments will centre on the question of a 'protocol' — the question of whether the members of the treaty should now

commit themselves to particular limits for the emission of greenhouse gases. The protocol to the Vienna Treaty on the Preservation of the Ozone Layer is every NGO's model in this connection; the continued production of chlorofluorocarbons has now been banned in most places under the Montreal Protocol. Why not sign a similar ordinance at Berlin? That will be the clamour.

The reasons why not are simple (see also D. G. Victor & J. E. Salt *Nature* **373**, 280–282; 1995). First, there are serious gaps in present knowledge of the sources of greenhouse gases. Not only in poor countries with other claims on their resources is the careful assessment of greenhouse-gas production regarded as a luxury; there are many European countries in which crude oil consumption (taken from customs records) is a proxy for carbon dioxide production. Second, there are problems with the minor (but not so minor) greenhouse gases, methane in particular. If rice paddy-fields are a principal source, does that mean that poor countries will have to cut back on rice production to allow rich countries to keep on driving automobiles? Then there is the problem of equity generally: should a country's quota of greenhouse gases be determined by its population (and when?), by its present production or by its expectations of economic growth? And what should be demanded of countries (Russia for example) that might have cause to welcome global warming? Should they forgo the use of fossil fuel, and the benefits its use may bring, in the interests of the rest of the world? These questions must be settled before anything like a protocol can be agreed.

Protocol

That is why the European Union's proposals for the meeting at Berlin make sense (see page 203). Let us have a protocol two years from now, the EU says, and let there be careful discussion in the meantime about the principles on which it should be based. Five years would be a safer estimate, but even that may not be enough. For what has not yet been appreciated is that a strict protocol accompanying the Rio treaty will be every bit as onerous a usurpation of national sovereignty as any of the arms control treaties signed in the past 50 years. In due course, there will be international inspectors trudging wherever on the surface of the Earth they choose, asking for measurements of the carbon dioxide output from this or that manufacturing plant, or the methane production from this or that paddy-field. That will be a more bitter pill to swallow than anybody has so far thought of foisting on the world at large.

The NGOs will be saying, in Berlin, that there is no time for leisured discussion of how the future should be managed. To the extent that the costs of abatement will rise with time, they will be correct. But it is also plain that, with the Rio treaty, the world embarked on a gigantic undertaking, nothing less than the detailed regulation of its fuel consumption and other important economic sectors, some not yet identified. It would be folly to set off on the wrong road in such delicate circumstances. And to do so without having the whole world on side would be dangerous for the reasons that unilateral nuclear disarmament is a fruitless policy □

House Republicans float proposal for single Department of Science

Washington. A proposal to establish a Department of Science to oversee most federal science and technology programmes is gaining ground in the US Congress, where several powerful Republicans think it could both save money and strengthen political support for science.

The idea of a science department has long been supported by the new chairman of the House of Representatives Science Committee, Robert Walker (Republican, Pennsylvania). It would also tie in with separate plans to abolish four existing government departments, including the Department of Energy. These plans were endorsed last week by Senator Robert Dole (Republican, Kansas), the majority leader in the Senate and a likely Republican candidate in next year's presidential election.

Walker and John Kasich (Republican, Ohio), chairman of the House Budget Committee, are trying to persuade those in favour of the abolition of the departments that creating a science department would be the best way to deal with the remnants of the Department of Energy, namely the energy research programme and the nuclear weapons programme.

According to Walker, the abolitionists had wanted to make the energy research programme part of the National Science Foundation. But at a recent National Space Society forum, he said that such a union would be a "mismatch". More importantly, there would be nowhere to put the nuclear weapons programme, except back in a

stand-alone agency such as the former Atomic Energy Commission — which, Walker says, would not yield any administrative cost savings.

Walker is proposing a science department that would incorporate the National Aeronautics and Space Administration, the National Science Foundation and the research arms of the Environmental Protection Agency and the energy and commerce departments. The National Institutes of Health would not be included.

The idea of a single government department controlling the bulk of research funds is feared by many scientists, who argue that US science has benefited immensely from the diversity of available funding sources. But advocates say that science would benefit from administrative savings, from having a cabinet-ranking science secretary and from a budget allocation which would be dedicated to science and therefore — the argument goes — less prone to attack in Congress.

Walker has not yet finalized his proposal. But he does expect to put forward a bill to establish a Department of Science during this session of Congress. The proposal is likely to meet opposition not only from the Clinton administration and parts of the scientific community but also from members of both houses on committees that would lose jurisdiction over the merged agencies.

Meanwhile Newt Gingrich, the leader of the House of Representatives, said last week that Walker also attended that government "should spend more in the long run on

R&D, and less on partnerships with industry". He argued that the federal budget was large enough to allow this to happen, even if it were cut back and balanced. "You'd think that with \$12 trillion to spend over the next seven years, you would be able to get a fair amount done," he said.

Gingrich said that he had just come from a meeting with 100 biotechnology chief executives that had left him convinced that the government should help high technology industry by cutting capital gains tax sharply.

Accused of 'bashing' science by proposing the closure of the US Geological Survey (USGS) and the National Biological Survey (NBS), Gingrich denied that there was any proposal to shut the USGS, and said Bruce Babbitt, the interior secretary, was to blame for the pressure on the NBS. Everyone in the western United States, he said, assumed the NBS was being set up in order to produce a database that the federal government would use "to take over the ranch".

Colin Macilwain & Tony Reichhardt

MIT chemist named to head spy agency

Washington. John Deutch, professor of chemistry at the Massachusetts Institute of Technology (MIT) and a senior research official in successive Democratic administrations, is likely to become the next director of the Central Intelligence Agency (CIA).

Deutch, who is currently deputy defence secretary with responsibility for research, was nominated as CIA director last week by President Bill Clinton. The previous nominee, Michael Carns, a retired air force general, withdrew in the face of trivial but potentially damaging allegations about the immigrant status of a family employee.

Before accepting the nomination, Belgian-born Deutch obtained agreement that if, as is considered likely, his nomination is confirmed by the Senate, the CIA director will once again become a full member of the cabinet. Friends of Deutch had previously suggested that he was worried that taking on the CIA job might jeopardize his chances of later being chosen as president of MIT.

At the Department of Defense, Deutch has tried to encourage closer cooperation between the Pentagon's vast research and development operations and the commercial sector. He was also an ardent protector of defence research in universities.

Deutch is no stranger to Washington, having been head of the Department of Energy's research office under the Carter administration.

Colin Macilwain

£5,000 light bulb – and it's broken!

London. **London's Science Museum has paid more than £5,000 (US\$8,100) for one of the earliest electrical light bulbs made by Thomas Edison (1847–1931).** According to Brian Bowers, senior curator of electrical engineering at the museum, a particularly significant feature of the bulb is the platinum clamps (see inset). These were Edison's solution to the problem of attaching the filament to the lead-in wires, and are only 7 mm long.

The light bulb is thought to have been made in 1879. It will be displayed with a similar lamp made in 1880 but with carbon attachments for the filament.

Christie's

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Carbon and platinum were the only materials available able to withstand the heat of the filament. F.G.

Gene therapy approval may be rocky road for industry

Washington. Viagene, the San Diego-based biotechnology company, last week became the first company to gain approval for moving to the next stage of gene therapy trials for an experimental therapy for AIDS.

But the way in which the Recombinant DNA Advisory Committee of the US National Institutes of Health (NIH) reached its decision shows that industry and the RAC may face a bumpy relationship as gene therapy moves down the clinical pipeline.

At present, the RAC has to approve any proposed gene therapy — whether carried out by industry or academic institutions — that either receives funds from the NIH, or is carried out at an institution that does so. In addition, the need for public acceptance has encouraged industry to seek RAC approval, even when this has not been legally required.

In moving to phase II trials, however, industry had hoped to deal primarily with the Food and Drug Administration (FDA) and local institutional review boards. Both have responsibilities for safety, and industry believes that this approach would save them time, as well as protect their confidentiality.

Indeed, Sheryl Osborne, director of regulatory affairs for Viagene, says that her company believed that because it had RAC approval for phase I of its protocol, it did not need to submit its proposals to RAC for phase II. It was on this basis that Viagene began late last year to recruit 190 patients at medical centres around the country.

Most of these centres do not receive funds from the NIH. Those that do said that they believed RAC approval was necessary, even though Viagene's protocol was not financed by the NIH. The RAC confirmed that view. As a result, phase II of the trial has gone ahead at centres not receiving NIH funds, while the others are awaiting RAC

approval to begin recruitment.

Osborne asked the Office of Recombinant DNA at the NIH to consider their protocol for expedited review. It agreed, and approval was granted last week. But she also wrote to Nelson Wivel, director of ORDA, arguing, that RAC review "should not extend into expanded protocols such as those for phase II and phase III."

Arguing that the safety criteria for inclusion in a phase II trial are very different to those for phase I, the RAC rejected the argument unanimously. This irritated but did not surprise industrialists. "The design of clinical trials is not easy," says Richard Daifuku, responsible for clinical trials at Targeted Genetics. "If you are a leader, you don't want to reveal your whole strategy if you don't have to."

During the same meeting, the RAC took its final vote to approve a simplified review process (consolidated review) for protocols that introduce no new ethical or scientific issues. Until now, all protocols have had to undergo a full review by both the FDA and the RAC. The FDA will now go ahead with its usual process, which requires a response within 30 days, while staff at ORDA with three reviewers from the RAC will decide whether the protocol needs full review.

The bureaucratic path to this change began last year when the National Task Force on AIDS Drugs Development identified the dual RAC/FDA process as a barrier to AIDS drug development. Leroy Walters, chairman of the RAC, says that the committee has finally found "a way to streamline the process for the investigator, while keeping public review." Alan Goldhammer, director of technical affairs for the Biotechnology Industries Organisation remains cautious. "That remains to be seen," he says.

Helen Gavaghan

Medical centre cuts staff in preparation for 'managed care'

San Francisco. The Medical Center at the University of California, Davis, announced last week that 82 employees are to lose their jobs next month in the first phase of a major restructuring aimed at limiting the impact of 'managed care' strategies now sweeping California's health-care system.

"We absolutely have to do this in order to keep our teaching and research abilities," says Robert Chason, chief operating officer of the Sacramento facility, who warned that some research programmes could be closed. "We can't continue to do everything."

According to Chason, another 90 vacancies will not be filled, and 300 temporary employees will not have their contracts renewed. The hospital employs 10,000 staff, and has an annual budget of \$368 million.

Chason says that the hospital is having to reorganize in order to attract patients and reduce costs as managed care plans overtake medical insurance throughout the state. Under managed care, primary care physicians become the "gatekeepers" for patients, directing referrals to specialists and turning to an advisory board to set treatment guidelines.

As networks of service providers begin to control patient care, Davis must link up with primary care physicians and create its own network to survive, says Chason. The medical centre has bought two primary care practices in nearby towns, and aims to add 100 primary care physicians over the next five years. Some of the new doctors are likely to be recruited from the Davis medical school, well known for primary care training.

The hospital is also planning affiliations with other hospitals and has established links to rural physicians extending to the Oregon border. Rural areas are largely overlooked by major hospitals, says Chason.

With federal and state dollars shrinking, the UC-Davis Medical Center has to compete more effectively with the private sector, he says. He estimated that the hospital must retain 175,000 patients within its system to be able to continue its teaching and research functions.

An advisory board that now links the medical school and hospital will be considering ways to integrate the two and set priorities. Physicians at the hospital have also been charged with reducing costs in a variety of ways, including reducing the use of laboratory and X-ray procedures.

University of California teaching hospitals in San Diego, San Francisco and Los Angeles are already considering similar changes. Chason predicts that medical schools throughout the country will face the same challenge as the managed care approach spreads.

Sally Lehrman

US sets up new bioethics advisory board

Washington. The US National Science and Technology Council was this week expected to approve plans for a National Biomedical Ethics Advisory Committee (NBAC).

The creation of such a body has the backing of both the Clinton administration and Senator Mark Hatfield (Republican, Oregon), chairman of the Senate Appropriations Committee. Officials expect the committee to be set up towards the end of the year. Its budget will be about \$2 million, taken from discretionary funds of a number of agencies.

The idea of establishing an ethics advisory committee originated shortly after President Bill Clinton took office,

prompted by congressional concern about the lack of policy for the handling genetic information. This will be one of the committee's priorities; another will be to evaluate issues arising from the need to obtain informed consent before carrying out research involving various groups of individuals such as the mentally ill.

The NBAC, which will include scientists, lawyers, philosophers and theologians, will be free to pursue any issue with a wide-ranging effect on public policy. For example, any unresolved details remaining after the disbandment of an advisory committee specifically set up to look at this issue are likely to pass to NBAC. H.G.

Europe agrees to seek new climate targets

London. Environment ministers of the 15 member states of the European Union (EU) have agreed to press for a new protocol on reducing carbon dioxide emissions beyond the year 2000 at the summit meeting on climate change in Berlin at the end of the month.

The decision was taken last Friday, in the run-up to the first meeting of signatories to the United Nations Framework Convention on Climate Change (FCCC) since the Earth Summit in Rio de Janeiro in 1992 at which the convention was signed. But there is still doubt whether EU states as a whole will fulfil their commitment to reducing greenhouse gases to their 1990 level.

The EU's decision is widely seen as indicating that it will try to take a lead in the Berlin negotiations, and will press for a mandate laying out the framework for the next series of commitments. These will probably be agreed at the third conference of parties in 1997.

But judging by discussions at the Intergovernmental Negotiating Committee in New York last month (see *Nature* 373, 462; 1995), EU states are likely to face a wide

range of alternative proposals put forward by other signatories to the convention.

The European environment ministers have yet to set any clear targets or time-scales for reductions after 2000. But they did list possible measures, such as controlling emissions from large combustion plants — the only major emitter not to be regulated in the United Kingdom, for example — the use of economic instruments, the promotion of new and renewable sources of energy and controls on transport.

Environmental groups say they are relieved that the proposals first agreed in December have not been watered down. But Matthew Spencer, atmosphere campaigner for Greenpeace UK, says that the main issue is "whether they will stick to their guns and actually push for the mandate, or cave in under pressure from countries such as the United States and Canada".

Total carbon dioxide emissions in the EU are expected to increase by between five and eight per cent over 1990 levels by the year 2000, even though some individual countries are expected to meet their target of zero increase. These countries, which include the United Kingdom, are pressing for others to meet their target and commit themselves to new targets beyond 2000.

The day before the environment ministers met in Brussels, the British government launched a report from the Department of Trade and Industry containing projections on UK energy use and the energy-related emissions of carbon dioxide between 1995 and 2020. Timothy Eggar, the energy minister, welcomed the report as the first to go beyond the year 2000, but admitted that "the further one looks to the future the less easy it is to predict."

But John Gummer, the Secretary of State for the Environment, said Britain would not only press for commitments beyond 2000 at Berlin, but would be seeking to limit the emissions of all greenhouse gases. "Negotiations must start in Berlin towards agreeing new commitments which, using a comprehensive or 'basket' approach, would aim at beginning to reduce total greenhouse gas emissions in developed countries below 1990 levels."

The UK delegation will argue that the more time governments are given to think about new targets, the better equipped they will be to anticipate the measures that need to be taken and thus reduce the risk of drastic cuts in the future.

Fiona Gammie

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Tighter emission controls suggested

Aid agency under fire in Brussels

London. Edith Cresson, the European research commissioner, is refusing to increase support for an intergovernmental organization set up to channel European aid to scientists in the former Soviet Union (FSU) until the heavily-criticized agency improves its efficiency.

The 'International Association for the Promotion of Cooperation with Scientists from the Independent States of the FSU' — known as INTAS — was set up in 1993 on the initiative of the then twelve member states of the European Union and five other European governments.

Backed, ironically, by Cresson's mentor as French prime minister, President François Mitterrand, the founding states promised to provide ECU50 million (US\$66 million) over two years to support joint projects involving scientists from the former Soviet Union (FSU) states working with colleagues in at least one European country.

By last summer, more than 500 projects had been approved in principle for funding. But many scientists have become increasingly frustrated at lengthy bureaucratic delays, in both Brussels and Moscow, which have resulted in some projects approved in 1993 still not being funded.

"It has been a total shambles," says Norman Dombey, of the department of physics at the University of Sussex. He was informed early in 1994 that a proposal for ECU23,000 involving physicists in Russia and Georgia had been approved, but says that none of

the Georgian scientists has yet received any money.

INTAS officials admit that it has taken longer than initially anticipated to make the necessary arrangements to transfer funds to Russia and other FSU states. They point out, for example, that the contracts need to be signed by all the participating scientists.

Nevertheless, they claim that there has been a "incredible" improvement over the past few months in the number of new projects that have been finalized.

So far, however, the Commission appears to be unimpressed. Keen to maintain the momentum behind the scientific aid programme, various EU member states had been pressing for an increase in the ECU5 million that the Commission is planning to provide to INTAS this year to support further joint research projects.

At a meeting of the EU Council of Research Ministers in Brussels last Friday, Cresson said she thought that INTAS was a good idea. But she drew the ministers' attention to what she described as the "dysfunctioning" of the organization which, she said, had so far signed "a very limited number of contracts" with FSU scientists.

Rejecting suggestions from some states for increasing support for INTAS, an independent agency under Belgian law, to at least ECU12 million, Cresson said that the Commission was not prepared to increase its funding unless the organization "puts its management in order".

David Dickson

EC 'likely to reopen patent negotiations'

Paris. The European Commission (EC) is under pressure from member states of the European Union (EU) to resurrect its attempts to develop a directive on patenting biotechnological inventions. This follows the rejection of a draft directive earlier this month by the European Parliament, concluding seven years of negotiations (see *Nature* 374, 103; 1995).

A Commission spokesman says that no formal decision has been made to propose a new version of the rejected directive. But one commission official says that, on the basis of discussions during an informal meeting of EU internal market ministers last weekend in the French resort of Biarritz, this outcome is "very likely". □

Prospective studies urged of health risks

Paris. Fundamental researchers have a social responsibility to help clinicians to anticipate risks associated with the use of therapeutics of human or animal origin, according to the report of a study on such risks carried out by the French biomedical research agency INSERM.

Focusing for example on growing concern about spongiform encephalopathies — which include Creutzfeldt-Jacob syndrome and bovine spongiform encephalopathy (BSE) or 'mad-cow' disease — it recommends that INSERM should create a group specialising in this area because of the "uncertainties surrounding the exact nature of these agents, their genetics and epidemiology, and the absence of diagnostic tests".

The report took two years to produce, and was prepared by a task force of 68 researchers from inside and outside INSERM on the basis of a literature search that produced more than 30,000 references. Their work was complemented by discussions with a smaller number of experts in the various fields covered by the report.

The report is the first major result of a 'collective expertise' service set up last year by the research agency to "short-circuit" the

traditional research timetable (see *Nature* **368**, 488; 1994). The service is intended to provide governments and companies with fast answers to contemporary questions of public health, and to analyse potential health hazards and opportunities, while simultaneously guiding INSERM's own research strategy.

The 45 chapters of the report cover topics ranging from kidney transplants to gene therapy, and blood transfusion to artificial reproduction.

Jean Rosa, the coordinator of the report and a researcher at the INSERM Laboratory of Molecular Genetics and Haematology at the Henri Mondor hospital in Paris, says that medical risks are increasingly due to the rapid progress in fundamental research. Another factor is the too-rapid application of such research because of increasing demand for new treatments.

As a result, the report concludes, fundamental research on the risks of new — and established — therapies needs to be increased. It also asserts that opportunities exist for more involvement in risk studies by fundamental researchers.

The report says researchers have a large

reservoir of knowledge and skills that is not sufficiently exploited by clinicians. New treatments should be subject to prospective studies of their potential risks, it argues, asserting that promoting such work is the responsibility of researchers towards society.

Even if the so-called prion diseases are rare, says the report, they are worrying because the infectious agent has not been unequivocally identified, and is "potentially present" in foodstuffs. Such infectious and neurological complications are important, it says, because of their seriousness, their relative frequency, the attention given to them by the media, the difficulty in evaluating the risk and the threat that the medical establishment will find itself under attack.

Complaining that France generally lacks an interest in prion diseases, the report says that such research as exists lacks coordination, with little connection to related work being carried out by the agricultural research organization INRA. It proposes more epidemiological research, and research on the transmission of prions and the development of diagnostics to detect healthy carriers and diseased animals.

Declan Butler

Lab rejects report that nuclear dump 'could explode'

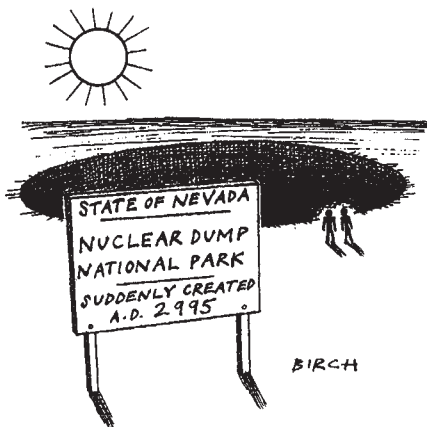
Washington. The Los Alamos National Laboratory in New Mexico has taken the unusual step of releasing internal peer review reports in a bid to quell public concern over suggestions by two of its scientists that a proposed underground nuclear waste repository could "go critical", triggering a chain of explosions.

The reports were extracted by Senator Bennett Johnston (Democrat, Louisiana), a supporter of the proposed repository at Yucca Mountain, Nevada, and read on the floor of the Senate in Washington last week. They say that the probability of the events predicted by the scientists, Charles Bowman and Francesco Venneri, is "essentially zero", and dismiss their work as being "of no technical merit".

In a paper prepared at Los Alamos, Bowman and Venneri had argued that plutonium and other fissile material stored in barrels underground would eventually leak, and that, mixed with rock that would act as a moderator, they could reach criticality and explode. A copy of the paper was obtained by *The New York Times*, which ran a front-page story under the headline "Scientists Fear Atomic Explosion of Buried Waste" in its issue of 5 March.

Bowman, a particle physicist, is the main sponsor of a nuclear waste disposal technology called accelerator-driven transmutation technology (ADTT), which he concedes is in competition with Yucca Mountain (see

Nature **370**, 404; 1994). "In the world that we live in, you look for the weakness in your competition and try to exploit it," he says, justifying the paper. Bowman's ADTT team has been pushing — without much success so far — for up to US\$500 million from the Department of Energy to build an ADTT demonstrator at Los Alamos.



Under its internal review process, Los Alamos set up 'red', 'blue' and 'white' teams of ten scientists each to act as prosecution, defender and judge respectively in assessing Bowman and Venneri's work. The white team finished its first assessment in December, and a reassessment of a modified paper last week. It accuses the authors of "showing

no grasp of elementary concepts" and making "alarmist estimates of potential effects, which have become less credible and more shrill throughout the process".

Some Senate staff members have been surprised that the Los Alamos laboratory made this damning assessment available to Johnson, but not to *The New York Times* when it was working on its story.

Managers at Los Alamos point out that peer review criticisms are not intended for publication, and suggest that the story got into print because Los Alamos scientists who discussed it with the newspaper "bent over backwards" to protect Bowman and Venneri. Bowman says he stands by his work that he hopes it will win support from another (unnamed) federal laboratory that is now reviewing it.

Most physicists outside Los Alamos dismiss the idea of spontaneous nuclear explosions in underground nuclear dumps for at least two reasons: that various chemical elements in rock will "poison" a chain reaction, and that, in the event of an energy release, there would be no containment mechanism to force an explosion.

But the story, which received wide television and radio coverage in the southwestern United States, is likely to linger in public perception, further undermining the energy department's already fraught efforts to store the bulk of US high-level nuclear waste at Yucca Mountain.

Colin Macilwain

Earthquake prediction 'likely to grow' after new review

Tokyo. Japan has launched a review of its massive programme of research into earthquake prediction, following the recent Kobe earthquake. The review is intended to make suggestions about how the programme might be strengthened.

But some seismologists involved in the review process fear that it will merely recommend that its budget should be expanded still further — without any change in direction — even though it has so far failed to predict any earthquakes.

The prediction programme, which entered its seventh successive five-year phase in April 1994, employs about 500 researchers throughout Japan. During this fiscal year, which ends on 31 March, it received ¥10.6 billion (US\$116 million) from government ministries and agencies.

In 1992 and 1993, the programme was subjected to a lengthy review process that included severe criticism from scientists outside earthquake prediction, such as the heads of seismological and volcanological societies of Japan. One issue raised was whether it was scientifically reasonable to assume earthquake prediction is possible.

But only minor adjustments were made as a result of this review, and the programme continues to focus on an empirical approach to collecting large amounts of data in the hope that they will eventually provide clues as to how earthquakes can be empirically predicted (see *Nature* **364**, 370; 1993).

According to some seismologists involved in the process, the present assessment also seems unlikely to lead to substantial changes, because it is being led by those responsible for the 1993 exercise. But it could help to attract more financial support for the programme from the government.

In normal circumstances, reviews of the programme are carried out about two years into each five-year phase. But after the Kobe earthquake, the Ministry of Education, Science and Culture asked the earthquake prediction subcommittee of the Geodetic Council of Japan — an advisory body of the ministry — to submit comments on the programme.

The subcommittee is chaired by Harumi Aoki of Nagoya University, and held its first meeting two weeks ago. It set up a drafting sub-subcommittee to review the programme and make recommendations for the future in the light of Kobe.

But to the dismay of those members of the subcommittee who have criticised the current direction of research, the drafting committee was hand-picked by Tomowo Hirasawa of Tohoku University, one of the senior leaders of the current earthquake prediction programme. Those who have

been pushing for reform were excluded.

The committee has already started circulating a document that is said not to contain any radical proposals for reform. The subcommittee was due to meet last Tuesday (14 March) to approve the document, which will be passed on to the parent Geodetic Council for approval in mid-April.

In the past, the council has simply approved the subcommittee's recommendations without change. Indeed, during the last review process one council member not engaged in earthquake research threatened to resign because the council was being asked to approve proposals it had not been involved in formulating.

One member of the earthquake prediction subcommittee fears that the whole review process is merely an excuse for the ministry to obtain more funds. He and other critics on the subcommittee have been calling for more emphasis on basic research into the mechanisms of earthquakes, rather than merely the empirical collection of data.

They also argue that earthquake researchers should be what they describe as more "honest" in telling the public about the present limitations of earthquake prediction techniques.

Another subcommittee member, Hiroshi Wakita of Tokyo University, who is a prominent researcher in earthquake prediction, describes the criticisms made by some scientists as "very valuable". But it remains to be seen how far they are taken into account in the review.

Meanwhile a special committee on earthquakes formed after the Kobe earthquake by the Liberal Democratic Party (LDP) — the most powerful party of the coalition government — has recommended that the government should look at the possibility of increased funding for what it describes as strengthening research into earthquake prediction, including monitoring the behaviour of animals, the level of well water and electrical currents in the Earth.

A member of the earthquake prediction subcommittee says he does not expect such recommendations to influence the research strategy of the programme because it is not the role of the LDP to suggest directions for research. But he says that the LDP committee could be "very influential" in winning extra funds for earthquake prediction.

Wakita does not expect final recommendations of the review process to emerge until about June. But this would still allow time for the various ministries and agencies involved in earthquake research to prepare budget requests for the fiscal year 1996, as these have to be submitted in late August.

David Swinbanks

UK seeks enhanced science links with South Korea

London. Clarification of European Union (EU) rules on applications for funding of joint projects between member and non-member countries is called for in a report on scientific collaboration between the United Kingdom and the Republic of Korea.

As the rules stand, third party participation is possible in some, but not all, EU-funded programmes such as information technology and biotechnology. But the project must have mutual benefit to both the EU and the third party, and non-EU member country researchers would have to pay their own share of the project.

"Roughly half of all Korean manufacturing investment in the EU is made in



Chung: need for open collaboration

Britain," said John Horam, Britain's new junior science minister. He welcomed the report and said links between the two countries had "grown steadily in recent years".

Historically, however, Korea has not looked to Britain for fostering scientific links. A recent survey highlighted in the report revealed that half of 360 Korean companies contacted preferred links with Japanese companies, 27 per cent with US companies and 11.1 per cent with German companies. France and Italy also came above the United Kingdom.

The report's aim, said Sir Geoffrey Allen, chairman of UK/Japan and Asia-Pacific Advisory Group, was to propose practical ways of increasing links between scientists in the two countries. Examples included establishing fellowships to enable UK postdoctoral researchers to work alongside leading industrial researchers in Korean companies.

Dr KunMo Chung, Korea's science minister, said that science needs more technology transfer and interdisciplinary activities. "We need really open international collaboration," he said.

Marking the publication of the report, Stewart Miller, the Rolls-Royce director of engineering and technology, signed a memorandum of understanding with Sang-Kee Suh, president of the Korea Institute of Machinery and Metals (KIMM), setting up a joint research centre at the Rolls-Royce Materials Research Centre in Derby.

On the same day, University College London opened a joint research laboratory between three of its departments and the Satellite Technology Research Centre at the Korea Advanced Institute of Science and Technology.

Fiona Gammie

France seeks to tighten its control over life sciences

Paris. An announcement last week by France's Ministry of Higher Education and Research that it intends to take direct control of a new national strategy for the life sciences suggests that the government may be on a collision course with the main research organizations over who should control the future direction of research.

Reinforcing this impression is the fact that, at a press conference held to announce the strategy, François Fillon, the minister for higher education and research, was flanked by civil servants — not by senior officials from the research organizations, a large number of whom sat among the audience.

Under the new strategy, the minister will set up 14 scientific committees. These will be responsible for carrying out the peer review and financing of projects in 14 'strategic' programmes divided into seven priority areas.

There will, for example, be three programmes in genetics, namely structural and functional analysis of genomes, human genetics, and genetics and environment. Similarly, 'physiopathological mechanisms' will include four areas: cardiovascular physiopathology and pharmacology; biopathology of prions; cellular imagery and mechanics, and neurobiopathology; and the development of functional imagery.

In addition, there will be two 'horizontal' programmes, one on bioinformatics and one on biotechnology. Total funding of FF250 million (US\$50 million) over two years will come from ministry funds.

According to Fillon, projects will be selected for funding if they are considered to have an "innovative dimension" that has been given inadequate attention by the national research organizations.

An official from the Centre National de la Recherche Scientifique (CNRS) says that he welcomes the fact that the science ministry has been able to obtain extra funds for life sciences from the finance ministry, particularly in the austere financial climate.

Of the promised FF250 million, FF153 million is guaranteed, while the remainder will be subject to approval by parliament in the autumn. These figures are significant, given that total funding of CNRS life science laboratories (excluding salaries) this year is FF346.4 million.

But the same official is concerned that the ministry has decided to administer the new money directly rather than pass it on to the research organizations. One explanation, say observers, is that the money was obtained by groups at the ministry that favour strong government intervention. Claude Griscelli, an adviser to Fillon, is considered the architect of the new plan.

Indeed, giving responsibility for the management of the programmes to committees within the ministry seems to fly in the face of Fillon's much vaunted policy of defining a national strategy by negotiating objectives (and their budgets) with the research organizations themselves through a new system of five-year contracts. "Does this mean Fillon lacks faith in his ability to administer such contracts?" asks one observer.

Others are sceptical that the ministry, which has only a small staff, most of whose time is tied up with university issues, has sufficient resources to manage 14 national programmes. They point out that research organizations such as the medical research agency (INSERM) already have well-developed systems for evaluating programmes.

Declan Butler

Gravity scores one, microgravity zero

Paris. Microgravity research crashed to Earth two weeks ago when bad weather forced a helicopter to jettison a Russian Photon 10 capsule containing a series of experiments in life sciences. The helicopter had been transporting the capsule from its landing site in Kazakhstan to the nearby airport at Erenbourg.

The capsule, which landed on 3 March after 15 days in orbit, was seriously damaged, and the accident also destroyed the French instrument IBIS (Instrument pour la Biologie dans l'Espace), which was flying the first of a series of planned biannual missions.

According to the French space agency CNES, 50 of the 64 cassettes containing biological material used in the IBIS experiments have been recovered. But it

will be several weeks before their condition is known.

The accident represents a "complete catastrophe", according to Lydie Gualandris-Parisot, a researcher at the CNRS laboratory of molecular and cellular mechanisms of embryogenesis in Toulouse, which had flown a series of experiments on the development of the nervous and muscular systems of the amphibian *Pleurodeles* aboard the fated mission.

The laboratory has recovered fewer than one-fifth of the experimental specimens, she says, and these are in poor condition; obtaining any meaningful results from the partial data will be "next to impossible". The only way to recover six years of work spent planning the experiments, she says, will be to re-fly the experiment. **D. B.**

EU commissioner sets sights on closer ties with industry

Paris. Edith Cresson, the research commissioner of the European Union (EU), last week announced plans to create joint task-forces with the European Commission's industry directorate (DG III) in five areas: vaccines against viral infections, educational software and multimedia; and the "cars, trains and aircraft of tomorrow".

Cresson was speaking at a meeting of the EU research ministers, the first at which she has represented the Commission since her appointment late last year. According to an official, the setting up of the task forces confirms the growing role of the industry directorate in deciding research policy.

But observers say that Cresson made less progress on gaining support for her controversial proposal to use the ECU700 million (US\$925 million) held in reserve for the next framework programme to give the programme a greater industrial flavour.

The Council also agreed that better coordination is needed between the EU's research policies and those of the member states. But it described as "premature" proposals from the Commission to include in the framework programme joint programmes involving some — but not all — member states (see *Nature* 371, 728; 1994).

Nor was it yet prepared to back programmes set up by several member states outside the framework programme — even though power to do so was given to the Commission under the Maastricht treaty. Coordination should be pursued instead, it said, through existing structures.

This refers to the programme committees and the Science and Technology Research Committee (CREST), a body made up of two members from each country and chaired by the Commission.

One Commission official points out that no role in promoting coordination was proposed at the meeting for the European Science and Technology Assembly (ESTA), an advisory body mainly made up of scientists which was created last year by Antonio Ruberti, Cresson's predecessor. One interpretation is that the assembly is now unlikely to be given a major say in policy decisions.

Attempts by the Council to transform CREST into a more powerful tool of its own interests met with mixed success (see *Nature* 370, 166; 1994). The meeting agreed that CREST should be strengthened and given more political influence, enabling it to play a bigger role in defining longer-term EU research strategies.

But the Council was divided about whether the committee should be chaired in rotation by member states — as several would like — or be chaired by the Commission, as at present. **D.B.**

Smithsonian heeds physicists' complaints

Washington. The Smithsonian Institution in Washington DC has promised to consult the visiting public, the American Chemical Society (ACS) and other interested parties before implementing changes to a controversial science exhibition at the National Museum of American History that were hammered out at a meeting with angry physicists last month.

The exhibition *Science in American Life* — a bold pastiche of images on matters such as the conflict between basic research and commercial pressure, the attitude of physicists to the atomic bomb, and the impetus behind the development of the birth control pill — opened last April, financed by a \$5.3 million grant from the ACS.

But it has come under fierce public attack from another scientific body, the American Physical Society (APS), which says that it exaggerates science's failures and trivializes its accomplishments (*Nature*, 373, 371; 1995). The APS did not participate in either planning or paying for the permanent exhibition. But it has been noisily attacking the exhibition since its opening.

The APS accuses the show of omitting important achievements such as the space programme and the invention of the transistor, of blaming science for environmental and other social problems, and of being "politically correct" in its emphasis on the under-representation of women and minorities in science.

Some historians involved in the show and outside museum curators have sprung to its defence. Jeff Sturchio, for example, a historian working for the pharmaceutical company Merck, which paid for 100,000 exhibition brochures, describes it as "one of the best shows that has ever been put on to raise serious questions about the role of science in American society".

But after a meeting with APS leaders and museum officials last month, Michael Heyman, secretary of the Smithsonian, told reporters that there would be "some changes which will render the exhibition a lot more balanced".

Heyman declined requests for an interview. But a spokesman for the Smithsonian promised that other groups would be consulted before the changes tentatively agreed at the meeting are implemented. The proposed changes are:

- Revising the opening feature of the exhibition, an audio tape portraying arguments between Ira Remsen, a nineteenth century chemist at John Hopkins University, and his commercially-minded colleague, Constantine Fahlberg. The APS thinks the sequence, intended to highlight the conflict between pure science and commerce, comes over as

an unseemly squabble. Heyman is said to agree with this;

- Brightening forbidding lighting and removing barbed-wire fences accompanying descriptions of the Manhattan Project to build the atomic bomb; these were intended to convey an impression of life at the remote and top-secret project sites;

- Changing descriptions of physicists' involvement in the politics of the atomic

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

In the late 1800s, Hugh Brown (left) saw physics teaching as a key to 'racial uplift'. From *Science in American Life*, at the National Museum of American History.

bomb, including an acknowledgement of Linus Pauling's role in establishing an international ban on atmospheric tests;

- Including displays directing visitors to related Smithsonian exhibitions, such as one on information technology and the nearby Air and Space Museum, which show the accomplishments of science.

These changes may not satisfy scientists who think that the whole show transmits a negative message. "Frankly, I don't think much can be done about the present exhibition without the infusion of new funds," says Kumar Patel, this year's president of the American Physical Society.

But Burt Richter, director of the Stanford Linear Accelerator Center, last year's APS president and leader of the team that met Heyman, is less strident in his criticism. He carried out some market research of his own at the museum. "The kids thought it was terrific, they didn't pick up any of the negative stuff," he reports. Nonetheless, Richter says the exhibition "needs a tune-up; they tried to do too many things, and didn't do any of them very well."

Some members of the original advisory panel to the exhibition say that their views were ignored when it was being prepared. "It was a gut-wrenching experience," says C. Marvin Lang, a chemistry professor at the University of Wisconsin, who recalls working with "social scientists and pseudo-scientists who had no idea how science worked".

The ACS, which will be consulted before any changes are made, is stepping gingerly around the issue; although paying for the show, it had no more editorial control than a corporation sponsoring an opera. Having

rejected a move by hostile board members to pay up but withdraw its name from the exhibition, the society has publicly supported the show. Indeed, some leading members privately deride the APS's criticisms as "sour grapes" from a body that refused to help finance the exhibition.

"There were many people involved with the show," says Ann Messmore, director of public affairs at the ACS. "Each one has a different opinion." Now that Heyman has spoken, the ACS says it would like to see some changes made. "The comments from the public are wonderful and the kids are excited and thrilled — but if you take a group of scientists to it, they are not going to be happy with every aspect," says Messmore.

The man at the centre of the controversy, Art Molella, chief curator of the exhibition, says the show was not meant for scientists. Their hostility, he says, comes from dashed expectations, namely that science museums and exhibits are commonly conceived as "boosters" for science and technology. He hopes it will rouse the interest of members of the public unimpressed by the science hagiography available elsewhere.

Alan Friedman, director of the New York Hall of Science, says the exhibition breaks new ground by telling the story of how public attitudes to science have changed. "Science exhibits have tended to be very promotional and self-congratulatory," says Friedman. "This is an excellent exhibition, showing how we are coming to understand the complex relationship between science and society. The public will come away feeling more confident that scientists are capable of being introspective about their work."

Richter denies that the APS insists that science should be portrayed heroically. "Of course, science can be misused," he says. "What's missing here is balance."

Also missing is a meeting of minds between scientists and historians. Social historians in particular tend to see their field in terms of conflict, whether between ideas, classes or individuals. The scientists who have attacked the show do not see their own history that way: they think the positive contributions that science, through technology, has made to twentieth century society ought to dominate a show called *Science in American Life*.

Furthermore, while most museum exhibits portray the scientist with all the dignity of the long-distance runner, this one has him as a player in an ice hockey game without a referee. No prizes for guessing which image schoolchildren prefer: several hundred thousand people are already estimated to have passed through the exhibition.

Colin Macilwain

Politics of climate change

SIR — I welcome Sonja A. Boehmer-Christiansen's recent article¹ about the politics of climate change, but her presentation is ill-founded.

Boehmer-Christiansen can have no evidence for her assertion that the scientists contributing to the work of the Intergovernmental Panel on Climate Change (IPCC) comprise a narrow group, whose members have advanced the issue of climate change research to raise funds for the support of their own research. Her statement that the IPCC 1990 Science Assessment was produced by Britain, with major assistance from a small group researching ozone science inside the National Aeronautics and Space Administration, is absurd and misleading.

The IPCC process of assessment is open and transparent. It is based on available scientific literature, a peer-review process of the analysis is essential, different views that are scientifically well founded should be described and scientists from both developed and developing countries participate². The basic reports, as well as the summaries for policy-makers, are written by teams of authors from many countries and reviewed by hundreds of scientists.

Boehmer-Christiansen ignores the obvious driving force, that scientists may wish to call attention to findings indicating that continued increase of greenhouse gases in the atmosphere might lead to significant changes of the climate on Earth. This has generally been done in a careful and well-balanced manner.

Nor is it correct to say that "the global warming has [for IPCC and me as its chairman] become the justification for a crusade against materialism and for a 'new organizing principle' — the preservation of the Earth". This does not address the *raison d'être* of the IPCC, to develop a method to provide useful scientific knowledge for the political process.

The IPCC has always emphasized the uncertainty of our knowledge and it is grossly unfair to describe its work as a "skilful exercise in scientific ambiguity". Uncertainty is a reality and it does not diminish risk.

Few, if any, critical remarks about the IPCC conclusions have been published in the peer-reviewed literature by scientists active in the field. But some scientists and lobbying groups with various political agendas have entered the scene. Seldom do these critics rely on widely accepted scientific analyses. Boehmer-Christiansen's comments demonstrate a surprising naivety about the complex process of making climate-change policy.

The growing debate about climate change is no surprise. In this context the IPCC has been accused of not being

objective in its work. Of course no single scientist can be completely objective, particularly about as complex an issue as that of human-induced climate change, but the collective work led by IPCC is generally much more reliable than other attempts to summarize scientific research results for the political process.

Bert Bolln

(Chairman, Intergovernmental Panel on Climate Change)

Kvarnsavägen 6, 18451 Österskär, Sweden

1. *Nature* **372**, 400–402 (1994).

2. *Nature* **368**, 94 (1994).

Distorted views

SIR — Your article on the European Southern Observatory (ESO) (*Nature* **363**, 551; 1995) gives a distorted view of the situation regarding ESO's dealings with the Chilean government.

I did not "pledge" to anything, I did not state any intention to "take the government to international court for breach of contract". Finally, ESO has not entered "a chain of appeals and counter appeals with the Chilean courts" while the "government has stood aside".

There are three issues. The first has to do with guaranteed observing time for Chilean astronomers and guarantees of labour rights for Chilean workers and has been largely resolved through an agreement between the Chilean government and ESO that is awaiting signature and ratification.

The second has to do with the question of the ownership of the land at the time the Chilean government donated it to ESO. Private parties in Chile have claimed ownership at that time. ESO considers this an internal issue for Chile, to be resolved by the Chilean government, and ESO has carefully refrained from entering into this legal controversy.

The third has to do with ESO's immunity of jurisdiction from Chilean courts. This immunity has been clearly recognized by the Chilean government while the other pending issues are resolved. This is important to ESO and to the worldwide astronomical community because stoppage of the work and consequent financial losses would damage the Very Large Telescope project which is in a very advanced stage of completion.

We are confident that through the continued help of the Chilean government an appropriate solution will be found that will permit ESO to continue its work in Chile on behalf of European and world-wide astronomy.

Riccardo Giacconi

European Southern Observatory,

Karl-Schwarzschild-Strasse 2,

D-85748 Garching bei München, Germany

The use of Taxol as a trademark

SIR — I find it, to say the least, ironic that a publication that has adopted the name "Nature" for a magazine about scientific research concerning "nature" should be contending that a trademark of another company is generic (*Nature* **373**, 370; 1995): I must assume that Mother Nature and the trademark examiners were "asleep" when you adopted the name "Nature" for your magazine. The aphorism about people who live in glass houses comes quickly to mind.

Pharmaceuticals are a highly regulated field. Trademarks for pharmaceuticals are subject to review by trademark offices and health authorities around the world. Generic names for pharmaceuticals are also subject to regulation by the World Health Organization (WHO), as well as local authorities such as the United States Adopted Names Council (USAN) and the British Pharmacopoeia. After review by numerous authorities in numerous countries, Taxol was approved for our use as a trademark for an anticancer preparation. The generic term approved by WHO, USAN and the British Pharmacopoeia for this product is 'paclitaxel.'

Our Taxol anticancer preparation is sold in more than 40 countries under the trademark Taxol and the trademark is now registered in nearly 70 countries. It is well-known as a trademark of Bristol-Myers Squibb throughout the oncology community, as is the generic name paclitaxel. The generic name paclitaxel has also become well-recognized in the research community; there are numerous scientific articles using the approved generic name paclitaxel. The fact that some earlier scientific articles used the term 'taxol' as a trivial name has no bearing on whether Taxol is a recognized trademark among oncologists, which it clearly is. To change the brand name of our product, as you suggest, would cause massive confusion and endanger the health and safety of oncology patients. That is a risk we will certainly not take because of a magazine editorial or indeed for any other reason.

I hope for your sake that Mother Nature doesn't wake up and notice your misappropriation of her name. She's been known to have quite a temper.

Stephen Chesnoff

Bristol-Myers Squibb Company,

345 Park Avenue,

New York, New York 10154-0037, USA

■ *Nature* has never prevented anybody from using the word nature as a common noun; the Bristol-Myers Squibb trademark would end a well-established usage. — Editor, *Nature*.

Is science policy in the doldrums?

John de la Mothe and Paul Dufour

Science policy is in the doldrums, having run out of intellectual steam and strong political commitment. It needs shaking up if it is to support and guide science and technology in a global, knowledge-based economy.

POLICY-MAKERS throughout the Organization for Economic Cooperation and Development (OECD) are faced with a serious dilemma. On the one hand, science, technology and innovation are clearly in the ascendancy, broadly seen as essential to the comparative advantage of nations and becoming mainstream activities generously supported by governments — to the tune of nearly \$150 billion a year. On the other hand, focused science and technology policy — the nest of government programmes and policies designed to support research, development and innovation — seems to be in the doldrums, having run out of intellectual steam and strong political commitment. Governments are struggling to maximize the returns to society from investing in research but have not demonstrated an understanding of how best to work with a changing scientific community. How did this situation arise and how might it be resolved?

One reason is that while science and technology have succeeded in moving from the periphery to the centre of government and policy, these creative enterprises have exposed themselves to an array of evaluative criteria not germane to science itself. Political, economic, distributive, military and geopolitical considerations have all become increasingly important in the formulation of science policy without regard to the real needs of science. It is now a question of 'science policy for what?' rather than 'science policy for science's sake'. Scientists need to understand this. Whether or not the United States decides to participate in CERN (European Laboratory for Particle Physics), Russia in the Space Station, Japan in the International Thermonuclear Experimental Reactor or any of the G-7 nations in national information infrastructure programmes (such as information highways) often has more to do with non-scientific strategic considerations than research priorities. The same is true for other countries. Moreover, in recent years, deficits and debts have greatly reduced the freedom that politicians have in public spending generally, with government budgets — determining social, military, industrial as well as science spending — now being capped, consolidated and reduced. As a result, political rhetoric designed to elicit funds through international cooperation in science and technology is increasing.

This erosion of funding should come as no surprise. The scientific community — whose members have come to view research grants as an entitlement rather than a privilege — has for decades promised the public and politicians far more than it could deliver. The 'endless frontier' of science has not managed to

trend to continue. Yet expansion of science was as much a result of the postwar baby boom and the chilled nerves of post-Sputnik policy hysteria as it was of any power inherent in human technical prowess. Indeed, it was no coincidence that public science policy arrived on the scene (in the guise of technology assessment, technology forecasting and so on) roughly when the public consequences of the fruits of research were being felt or perceived. It was a growth curve that just had to flatten. Otherwise, as the sociologist Derek de Solla Price noted in 1969, if science budgets and science enrolments did indeed keep going up, by 2001 "every man, woman, child and dog would be a scientist". Or, as D. Allan Bromley, a former US presidential science adviser, recently argued, the past 40 years has been an unnatural blip, and scientists will have to get used to that. Science funding has entered a protracted steady state.

A third reason is that science policy has stopped being about science and is now principally about economic growth, job creation and industrial development. The focus on the development of conditions for the knowledge-based economy — which most OECD governments are desperate to develop (though not to define) — has brought a flurry of attention by (mostly) governments to the content and orientation of trade, investment, industrial and research activities. Thus, interest in a country's technological balance of payments, GERD/GDP ratio (gross expenditure on research and development as a percentage of gross domestic product), foreign direct investment, tax credits for research, the creation of clusters of small high-tech companies and so on has mesmerized policy-makers into a narrow quest for the economic indicator that will reveal their techno-economic future. This has led to a plethora of new programmes in which university researchers work with companies, participate in networks and collaborate with researchers in other disciplines. And to get funding, researchers concoct an array of economic benefits that might flow from their efforts. These range from new jobs for both researchers and nonre-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Doing more with less: inside CERN's LEP tunnel.

translate itself into an 'endless solution'. Science has therefore become a target for mainstream spending cuts along with every other government portfolio (see figure over page). The effect on science policy is that control on spending in science and technology is passing from the hands of research administrators to bureaucrats in departments of finance, commerce and industry. With it comes the trend of setting priorities and developing so-called 'performance measures' for innovative work.

A second reason for politicians' ennui with science policy is that the rapid growth of science budgets from the late 1950s until the mid-1980s was an anomaly, not the norm. A whole generation of young scientists trained in the aftermath of the Second World War expected this upward

CERN Photo

searchers to the shoring-up of failing companies and industries, as well as the creation of new businesses.

Implicit in much of this is a confusion within bureaucracies about the extent to which governments should intervene in the workings of the economy. One type of policy economist sees economic growth as a function of company-level activities in which the government has little or no role except to provide a stable fiscal environment. In this world, grants for research distort the market and artificially boost the salaries of researchers. What is more, science and technology policy — however defined — means to these economists a form of 'picking winners' (something that most people agree governments are ill-equipped to do and which is anathema to the private-sector free-trade mantra). In this model, governments are often seen as surrogates for industrial research and development. A second type of policy economist, however, agrees on the importance of companies but also sees cooperation between government and business as a key to successful competition in a global economy. In this particular world, science policy is not just about grants but is also about 'strategic trade', 'strategic technologies' and attracting investment into research and development. But for this type of policy view to work, a government must know where its science-based strategic interests lie. Many countries, however, have no clear idea about what these are. And scientists haven't helped. At one level, this might be easier for the United States, but it is far from obvious for those in countries such as Sweden, the United Kingdom, the Netherlands, Canada or Australia. At another level, the whole character of 'strategic interests' is being utterly re-defined in the context of the post-Cold War world. Scientists must help in this definition.

A fourth reason is that some governments do not know how to delineate clearly their goals for science and technology or how to orchestrate their activities to achieve the most effective use of public money. They may also lack the political will to reallocate public funds across the science portfolios. Although priority-setting is *de rigueur*, it is rarely accompanied by action plans and recommendations that can be implemented (eloquent White Papers and other policy statements notwithstanding). Given that billions of research dollars are spent annually by each of the OECD governments, it is hardly surprising that research administrators act less in the public interest than as members of a tribe, protecting their own budgets and programme mandates much as they would if they were protecting territory.

Science policy is not just about funding and delivering the goods, and science-policy ideologues should not be involved

in managing science budgets. Effective science policy is possible only when governments organize themselves in ways that allow them to receive sound advice on their science and technology activities. Every country has at some time benefited from (and been damaged by) charismatic advisers who gain access to the corridors of power by virtue of their political contacts and experience. Such people will always be available and consulted. But they are not sufficient. Questions need to be asked about whether a government can really afford to be without an independent chief science adviser, an office of science

held responsible for next year's products or the next quarter's bottom-line. Nor can the institutions of research be expected to change overnight to conduct missions for which they were never designed.

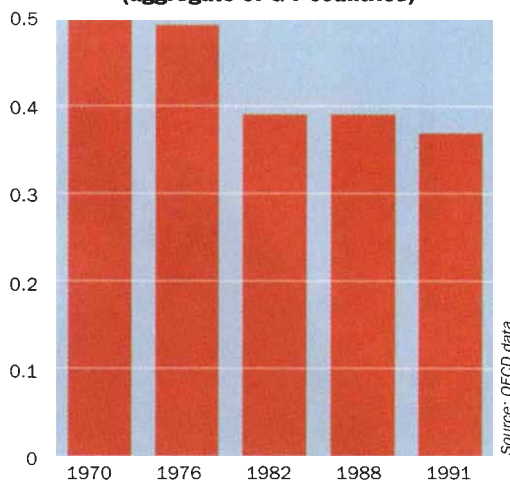
If science policy is to emerge from its current doldrums, it must focus on what it does best. In the 1950s and 1960s, science policy helped to build impressive science infrastructures throughout the OECD. In the 1970s and 1980s, science policy helped democracies to come to grips with the public impact of technology and assisted in the development of national competitive advantages. Today, science should not fall prey to industrial or trade policy theorists. Instead, it should extend its responsibilities to serving the global common ground and solving social ills, while maintaining its obligation to the advancement of knowledge.

The practitioners of science and the managers of science policy have some major challenges before them. First, they must learn to develop a stronger and clearer partnership with government in developing the future direction of research and science policy. This partnership must recognize the changing nature of the scientific enterprise itself, where innovation and scientific research are increasingly interrelated. Second, the research enterprise (and there is more than one) must become more sophisticated in

delivering its messages, priorities and needs to the public policy process. Scientific practitioners can indeed provide solid advice in the shaping of national policy for research, but this must be done in a way that is realistic to the task at hand. The need for performance measures, indicators and state of research policy/foresight reports will increase as governments seek to demonstrate accountability and transparency to the public. The scientific community can help in these efforts. Finally, the public dimension of the interface between science and government will be tested with new pressures, as we are seeing with issues related to ozone depletion, persistent organic pollutants, new reproductive technologies and research problems associated with ageing. The ability to respond to new demands from more knowledge-thirsty societies and truly global issues will strain the bond between national politicians and international scientists. Good science policy is needed as never before. It must be shaken out of its doldrums. □

John de la Mothe and Paul Dufour are with the Program of Research on International Management and Economy, Faculty of Administration, University of Ottawa, Ottawa, Canada K1N 6N5.

**Public funding of R&D
(aggregate of G-7 countries)**



Ratio of public funded R&D to GERD (estimate for 1991).

and technology policy, an office of technology assessment, an effective national academy of sorts, or an independent science-policy research unit. At the moment, only a few OECD countries are puzzling over these questions. Pluralistic structures for policy advice are not only more democratic; they can also help to shape effective policy. Governments should be asking whether it would be worth their while training their own interdisciplinary teams of people competent in the evaluation of research. And science communities must ask themselves whether it is good enough to act or to be perceived as acting as a mere lobby group to government for physics, biology or engineering, or for a space station or an information highway — rather than in the broad public interest.

Science and technology are serious activities. They must not be diminished by accountants' pens or by politicians seeking unrealistic goals. Nor should they be distorted by scientists' lurid claims or promises. Science and technology are important to advanced economies, but they clearly cannot replace the entrepreneur, the shrewd financier or the crafty mechanic. Science and technology are about creativity and insight, so all they can do is inspire and help to provide the tools of capitalist development. They cannot be

Celebration of QED's half-century

A badly constructed yet fascinating account of the origins of quantum electrodynamics suggests the need for a way of capturing both the spirit of an heroic age and the later history of the concepts that it spawned.

THIS year is thick with fiftieth anniversaries, mostly because the European phase of the Second World War ended 50 years ago. But 1945 also marked the return to academic laboratories of people who had spent the war at places such as Los Alamos, the Radiation Laboratory at MIT and its outposts (most significantly that at Columbia in New York), not to mention Chalk River and Britain's equivalent of the RadLab at Malvern.

One of those to return from the war was the late Maurice Pryce, one of the first of Oxford's professors to be appointed then. With his formidable reputation in theoretical physics, his knack for dressing well and his marriage to Max Born's daughter Margaret, he was a focus of attention.

Pryce's first course on quantum mechanics, begun in October 1946, was therefore filled with the curious as well as the eager. In the event, the course was an intelligent scamper through Dirac's great book *The Principles of Quantum Mechanics*. The importance of the inside information that Pryce kept pressing on his audience went largely unrecognized; he was full of exhortations to pay more attention to the energy levels of the hydrogen atom. He mentioned a fellow called Lamb, then at Columbia. Few appreciated that he was pointing to an inside track to fame and fortune.

The hydrogen problem was the springboard on which the great revolution of quantum electrodynamics was mounted. What follows is mostly about a remarkable book on that formative experience for the research community, *QED and the Men Who Made It*, by Silvan S. Schweber, a physicist turned historian who has gone back to physics to celebrate QED (Princeton University Press, 1994, \$99.50 (hbk), \$39.50 (pbk)).

Predictably, the names cited on the cover are Dyson, Feynman, Schwinger and Tomonaga (in alphabetical order), but Schweber casts his net wider than that. Thus he staunchly defends the value of the contributions made in the 1920s by Pascual Jordan, one of the authors (with Born and Heisenberg) of the famous and monumental article still called the *Dreimännerarbeit* (*Zeit. Phys.* **35**, 557–615; 1926).

Jordan seems to have contributed to that landmark paper the final section in which a vibrating string was a one-dimensional analogue of the radiation field; each vibrational mode becomes the equivalent of one of the oscillator modes into which the radiation field can be decomposed. Schweber

guesses that Jordan's contribution would now be more honoured had he not become a member of the Nazi party and, separately, taken to sports such as psychology, all the while hampered in communicating with others by a crippling stammer.

If that seems trite, the same is also true of the biographical sketches with which Schweber's book is loaded. Was Schwinger's notorious habit of working only at night a simple consequence of his "shyness", and his fear of being "dominated" by pushy individuals? Schwinger (now, sadly, dead) would probably have a more interesting tale to tell. The anecdotal information that Schweber has painstakingly gathered is in many ways a distraction.

It is, of course, interesting and important to know what people such as Rabi and Oppenheimer did to create the intellectual climate in which the problems in electrodynamics were solved not once, but twice (by Schwinger and Feynman). But Schweber is a laundry-list biographer; if he has a piece of information, he puts it in, often by quoting at length from original documents. (There is also a laundry-list of a kind — the budget for a conference at Shelter Island, off the north coast of Long Island, in the spring of 1947, estimated to cost \$3,150.) His discussion of how the habit of solving practical problems for the war changed the temperament of physics in the United States is similarly unconvincing. The result is a book (at 732 pages) too long by half.

But one should not be too harsh. Although Schweber deals with physics in much the same way, reproducing passages from crucial articles and whole letters from the likes of Pauli and Dirac, the result is a kind of source-book, even a scrapbook, for those who want to know more about the revolution that set the trend and the pace for the breathtaking development of the rarefied end of theoretical physics in the past half-century. And Schweber, like Pryce all those years ago, is right to emphasize the importance of what Lamb did.

The issue, half a century ago, was the suspected difference of energy between the two possible configurations of the first excited state of hydrogen. An electron in the ground state must have zero angular momentum (and its wavefunction must be spherically symmetrical), but the first excited state (with principal quantum number 2) allows states *S* and *P* in which the angular momentum is either 0 or 1 respectively, and called $2S^{1/2}$ and $2P^{1/2}$. Simple wave mechanics shows the two states to be

degenerate. Lamb's experiment, a *tour de force* for the time, demonstrated an energy difference of about 1,000 MHz, with the *S* state at the higher energy.

Remarkably, at the same laboratory at Columbia, at the same time, Rabi and his associates uncovered another crucial piece of evidence: the magnetic moment of the electron is not exactly what it should be (1 Bohr magneton, the magnetic moment associated with one unit of angular momentum in an electron orbit), but some 0.12 per cent greater. These were the data with which the score of theoreticians gathered for the Shelter Island conference in 1947 were challenged.

Schweber's anecdotal method works to his advantage in reconstructing that event, largely with the help of notes kept by bystander participants. The penny seems to have dropped quickly. In each case, the explanation must be the interaction between the electron and the radiation field. The book tells how Hans Bethe made a non-relativistic calculation of the Lamb shift on the train to Schenectady, bound for a few days consulting for General Electric.

Instead of following the then standard practice of ignoring the infinite quantities notoriously obtained in calculations such as this, Bethe followed the advice of Kramers (from Leyden, who had been at Shelter Island) that the effects of the radiation field on the apparent mass of the electron should be counted as a part of its observed mass. That was the first application of the ungainly techniques of renormalization by which field theories are still beset.

From there, it is a helter-skelter tale. Half a dozen groups were busy with relativistic calculations of the Lamb shift; the unluckiest were Weisskopf and French, at MIT, who got the correct result but delayed publication because others told them of discordant conclusions. (There is a moral there.) Schwinger produced a general theory, Feynman developed his other scheme and Dyson proved that the two views were equivalent.

The three years from 1945 to 1948 set the style for everything that has followed since in the understanding of matter, radiation and the link between them. It is an heroic episode that amply deserves celebration. But Schweber could at least have guided his readers towards the uses afterwards made of quantum field theory in the understanding of other components of matter than electrons. That is also an heroic tale.

John Maddox

Destructive debris

Alan W. Harris

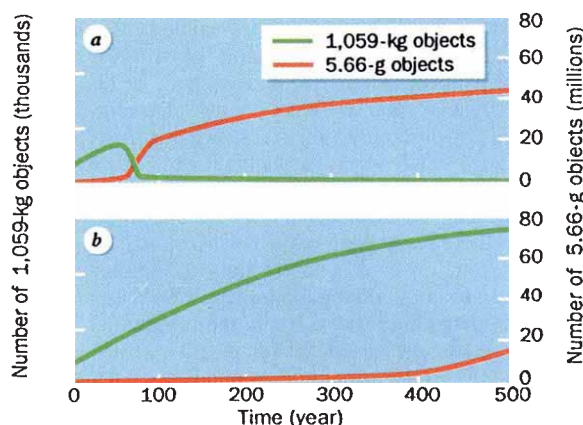
ARE we facing a 'cosmic catastrophe' on our own doorstep? Researchers in Pisa, Italy, calculate that collisions between payloads, rocket bodies and bits of debris in orbit about 1,000 kilometres above the Earth may lead to that region becoming a veritable shooting gallery within a century or so. Reporting in the *Journal of Geophysical Research*, Rossi *et al.*¹ make new use of a model of erosion and catastrophic breakup that was first developed to study the collisional evolution of the asteroid belt²⁻⁴. They find that a runaway in the population of centimetre-sized debris may occur in about 50 years. Once this runaway occurs, any newly introduced object of a metre or more across can be expected to suffer a disabling collision with such a debris particle within a few decades. Clearly such collisions would become a severe problem for unmanned spacecraft, and could make manned operations too hazardous to contemplate.

The realization of this problem is not new. Kessler and Cour-Palais⁵ discussed the possibility of the creation of a debris belt around the Earth in 1978, and it has since been a cause of concern to various agencies in the United States and Europe⁶⁻⁸. The Air Force Space Command, which routinely tracks orbiting spacecraft and large debris pieces, is currently conducting a survey of debris pieces down to a few centimetres in size, to assess the population of the small particles that are deemed most hazardous.

The key result of collision studies, whether of asteroids or of Earth-orbiting debris, is that the distribution of numbers of objects with size relaxes to a quasi-equilibrium which is fairly well represented by a power-law distribution, $dN(m) \propto m^{-q} dm$, where $N(m)$ is the number of bodies larger than mass m . Essentially, large bodies are broken up by collisions with much smaller bodies, thus supplying even more small bodies — and so the destruction continues. Small bodies are removed by collisions with even smaller objects, eventually breaking down to dust which is removed by solar radiation pressure or air drag.

The equilibrium value of the exponent q seems to be very stable under a wide range of assumptions, falling in the range $5/3$ to 2.0 . The value $q = 2.0$ corresponds to a distribution in which the total mass of particles in any logarithmic interval of size is constant; for instance, there is as much

mass contained in the mass range 1–10 g as in the range 1–10 kg. This is the transition between most mass in the largest particles ($q < 2$) and most mass in the smallest particles ($q > 2$). The value $q = 5/3$ corresponds to an equal-area distribution, where the total surface area of particles is equal in equal logarithmic intervals. If the size ratio of the threshold for a catastrophic breakup is constant with respect to size, then this corresponds to equal collisional lifetime for any particle size. The



Collisions above the Earth. These are the projected populations of objects in an orbital shell between 900 and 1,000 km up, according to the models of Rossi *et al.*¹. The two mass bins plotted correspond to the most prevalent mass of satellites and rocket casings launched, and to the size of debris particles capable of destroying such satellites according to their nominal model. *a*, Results for the nominal model, in which the specific energy for catastrophic breakup, Q^* , is 10^3 J kg^{-1} . *b*, Results for objects of the same size, but for $Q^* = 5 \times 10^4 \text{ J kg}^{-1}$. The runaway growth of the debris belt is postponed from ~ 50 years to ~ 500 years.

main belt of asteroids in the Solar System follows a number distribution close to $q = 11/6$ over many decades of mass⁹.

So what do Rossi *et al.* find? In the figure, adapted from their paper, panel *a* is a plot of how the population of two sizes of bodies changes with time, in one shell of orbital altitude, according to their nominal model. I have extracted the plots for mass bins centred on 1,059 kg and 5.66 g. The former is the most prevalent mass range of satellites and rocket bodies injected into orbit, and the latter is the characteristic size of debris capable of destroying such satellites. They assume initial populations of 869 small and 282 large objects, and a launch rate of 9.5 per year into the large mass bin. We can see that the population of the small debris particles increases nearly exponentially to more than 10^7 in about 50 years, leading to a catastrophic increase in the rate of destruction of larger bodies, to the extent that the equilibrium population of the latter drops to a very low

number, in spite of new launches.

Rossi *et al.* explore the sensitivity of their result to variations of the parameters assumed. They find that reducing the starting number and rate of launch of 1-ton bodies by about half each only delays the runaway by about 20 years. They point out that the model parameter that most affects the result is the collision strength of bodies. This is parametrized in terms of the specific energy required for catastrophic disruption of the target body, Q^* . The nominal value they adopt is 10^3 J kg^{-1} , which is derived from laboratory experiments of hypervelocity collisions with rock targets. At the velocities of typical collisions, around 10 km s^{-1} , this corresponds to a catastrophic breakup of a 1-ton spacecraft by a debris particle of about 20 g.

But orbiting spacecraft are not really very like rocky targets: they tend to be nearly hollow (at least in the larger sizes), and they are mostly made of metal, which is stronger and more ductile than rock. One can easily imagine the differences in collisional outcome by thinking of shooting a bullet at a glass bottle or metal can. Clearly the can is more resistant to total disruption. Carrying the analogy a bit further, it can be imagined that the ultimate result after many impacts is the same: the can is eventually reduced to fragments, but it takes more or larger impacts. Rossi *et al.* examine this by assuming a larger disruption energy density, up to $Q^* = 5 \times 10^4 \text{ J kg}^{-1}$, which corresponds to requiring a debris body of $\sim 1 \text{ kg}$ to disrupt a 1-ton spacecraft. The result of this model is summarized in part *b* of the figure. In this case,

catastrophe does not occur for about 500 years. It is noteworthy that the main change is the timescale to reach the equilibrium population of small particles (500 rather than 50 years). At equilibrium, the population of particles is about the same, so the collision timescale is about the same, with little dependence on assumed strength. That population is tens of millions of particles each about 10 g in mass, implying a collision timescale of a few decades between a 1-ton satellite and a 10-g debris particle.

The authors also turn their attention to possible ways of avoiding the predicted runaway in debris population. Even in the most extreme case of all launches ceasing forthwith, they find that it may already be too late to avoid the runaway. The above population of several tens of millions of 10-g particles amounts to a total mass of several hundred tons. We have already launched that much mass into long-term stable orbits at a height of $\sim 1,000 \text{ km}$. It seems, then, that it is only a matter of time

before mutual collisions reduce those bodies to 10-g pieces of debris. The question is, will this take decades or centuries? Over a period of time several times longer than that required to create the runaway, the cloud will naturally clear, as still smaller bodies erode away the offending size of debris. But this can only happen in the absence of the larger bodies that supply the cloud. That is, space would become safe again only if we quit going there — clearly not a welcome solution.

Does near-Earth space become completely unusable in the presence of a fully developed cloud of debris? For typical model simulations, Rossi *et al.* estimate the collisional lifetime of an individual spacecraft to be of the order of decades after the runaway occurs (although this is the timescale for a collision that breaks up the spacecraft; the timescale for one that would render it dysfunctional may be less). As typical spacecraft have functional lifetimes of only a few years, it may be that the risk of collisions will not be a great factor in the overall failure rate of robotic spacecraft. The prospect for manned spacecraft is more daunting. But collisions

in space may not be the greatest risk faced by astronauts. The current death rate from flying in space is of the order of one per 10 person-years spent in space, so it could be that flying in a fully developed debris cloud for weeks or months is no more dangerous than the risk associated with launch and return. □

Alan W. Harris is in the Earth and Space Sciences Division, Jet Propulsion Laboratory, California Institute of Technology, MS 183-501, 4800 Oak Grove Drive, Pasadena, California 91109, USA.

1. Rossi, A., Cordelli, A., Farinella, P. & Anselmo, L. *J. geophys. Res.* **99**, 23195–23210 (1994).
2. Wetherill, G. W. *J. geophys. Res.* **72**, 2429–2444 (1967).
3. Fugiwara, A. *et al.* in *Asteroids II* (ed. Binzel, R.) 240–265 (Univ. Ariz. Press, Tucson, 1989).
4. Petit, J.-M. & Farinella, P. *Celestial Mech.* **57**, 1–28 (1993).
5. Kessler, D. J. & Cour-Palais, B. G. *J. geophys. Res.* **83**, 2637–2646 (1978).
6. European Space Agency, Space Debris Working Group *Space Debris* (ESA Spec. Publ. 1109, 1988).
7. Interagency Group (Space) *Report on Orbital Debris for National Security Council* (Washington DC, 1989).
8. Office of Technology Assessment *Orbital Debris: A Space Environmental Problem* (US Govt Printing Office, Washington DC, 1990).
9. Dohnanyi, J. S. *J. geophys. Res.* **74**, 2531–2554 (1969).

FISHERIES

The primary requirements

John Beddington

CONCERN that fish are being harvested unsustainably has usually centred on the dramatic collapses of individual stocks or on the conflicts associated with the over-capacity of the fishing industry. At a global level, marked declines in the rate of increase of the total world catch, and indeed falls in that total, have recently prompted the question of whether catches are close to or exceeding the maximum that are sustainable.

On page 255 of this issue¹ Pauly and Christensen address the issue by calculating the level of primary production necessary to sustain world fish catches. They estimate that some 8% of global aquatic primary productivity is required, a figure that is much higher than previous estimates² but, on the face of it, not particularly large. However, when the contribution of different ecosystems is examined, major differences emerge. For the open ocean, the primary productivity requirement is only around 2%, but open-ocean fisheries contribute only a small proportion of the world fish catch. For the areas of greatest importance to fish production, for instance the continental shelves, the requirements range from 24 to 35%. Such figures imply that current levels of fishing — and certainly any increases — are likely to result in substantial changes in the ecosystems involved.

Pauly and Christensen's calculations

start from data collected by the UN Food and Agriculture Organization (FAO) and involve classifying catches of different species into different groups and calculating a fractional trophic level for each group: this statistic is essentially a weighted average of the levels at which species feed. For example, Antarctic krill are

given a fractional trophic level of 2.2 on the assumption that they feed 80% at the phytoplankton level and 20% at the herbivorous zooplankton level.

The energy transfer efficiency between trophic levels in aquatic ecosystems is commonly assumed to be 10%. Pauly and Christensen document estimates of transfer efficiency derived from a large number of models which show a mean of around that figure, which they therefore adopt. The energy requirement to sustain fish catches can then simply be compared to estimates of primary productivity for different ecosystems. The figures calculated are subject to a number of queries; as the authors point out, estimates of primary productivity are on the low side, but by contrast catches by artisanal and subsistence fishermen are probably under-reported. Nonetheless, such considerations are unlikely to alter the conclusions of the paper a great deal.

A notable complication that Pauly and Christensen take into account is the amount of fish caught and discarded for economic reasons. The annual level of discards has been estimated to be some 27 million tonnes³, a staggering figure in the context of a total world catch of less than 100 million tonnes.

One potential snag is that the analyses critically depend on the catch composition of the fisheries, and in particular on their trophic level. The authors avoided this problem by calculating average catches from the period 1989–91, when the species composition of the world catch did not alter markedly. However, there is a more important dimension to this question. If the level of fishing means that too high a proportion of primary productivity is being required, then it is likely that the ecosystems will move towards a composition increasingly dominated by lower trophic levels. This will tend to lower the primary production requirements.

It is well known that sustainable yield from aquatic ecosystems can be increased by harvesting species of lower trophic levels — there is the obvious gain to be made from saving in energy transfer; and species at lower trophic levels typically have somewhat higher ratios of productivity to biomass, and hence higher sustainable yields. Such a change is a probable consequence of fishing practice in any case. The shift in emphasis towards lower trophic levels can be intentional, or a natural consequence of overfishing of species at higher trophic levels. The first situation is likely to occur when the economically preferred species is at a lower trophic level (for example, shrimp). The second, and probably more common, situation occurs when the preferred species is from a higher level (for example, cod).

Pauly and Christensen's estimates provide a snapshot of how things stood at the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Up for grabs — pressures on fisheries stem from their open-access nature.

turn of the 1980s. Major changes in catch and species composition in aquatic ecosystems have certainly occurred in response to fishing. They are likely to continue, but whether primary productivity requirements will increase or decrease as a result is unclear. What is clear is that an unlimited movement down the trophic web in target species cannot continue without causing economic difficulties, losses in biodiversity and conservation problems for higher species in the ecosystems, including marine mammals. The importance of the work is to draw attention to these limits and to highlight the necessity of monitoring the relevant statistics at a global level.

The limit to which complexes of interacting species may be exploited is still an open question. Ignorance on that score is less of a drawback than it might appear, as the scientific issues of assessing the sustainable yield of individual fish stocks are relatively well understood. The limit to which a single species can be exploited as a proportion of its biomass is known to be a simple function of the parameters of growth and mortality. When considering harvesting at different levels in the ecosystem, it has been shown that, for a wide range of life-history characteristics, yield is proportional to natural mortality⁴, which itself tends to be inversely related to trophic level. Sustainable harvesting of fluctuating stocks can be addressed by feedback policies, and in principle it should be possible to manage fisheries sustainably.

Principle is one thing, however, practice another. The main force driving the level of exploitation to the unsustainable is overcapitalization of the fishing industry, which itself has been produced by the open-access nature of fisheries. The FAO has estimated⁵ annual losses of the global fleet to be some US\$54 billion. These losses are largely met by subsidies. The failure of fisheries management to address the problems of unrestricted access and to stop unsustainable levels of harvesting (which have led to the depletion of many major stocks) is estimated to be costing a further US\$25 billion each year. These figures show the potential benefits to be had from successful fisheries management on the world scale. □

John Beddington is in the Centre for Environmental Technology, Imperial College of Science, Technology and Medicine, 48 Prince's Gardens, London SW7 2PE, UK.

Inverse agonists exposed

J. W. Black and N. P. Shankley

ON page 272 of this issue¹, Bond and his colleagues claim to have come up with data "providing evidence supporting the existence of inverse agonists and validating the two-state model of G-protein-coupled receptor activation". The concept of inverse agonists and the related explanatory framework are not new. So far, however, they have not been widely accepted, perhaps because they did not seem relevant to most pharmacological and clinical experience. But the new data mean that the prevailing concepts of ligand actions need to be re-examined. The practical interest here is the effect that such a change in thinking might have on drug development.

Agonists (substances which are hormone-like in their ability to induce selective physiological responses in tissues) and receptors (the cellular sites where agonists act to express their selective effects) have been fundamental concepts in pharmacology for a long time. Over the years, both have evolved intermittently to accommodate observational and technical advances.

To account for the mutual antagonism (*sic*) between pilocarpine and atropine, Langley² assumed that they formed compounds with a receptive substance (receptors) in tissues in proportion to their "relative mass and chemical affinity". He based his concept on the "analogous case with inorganic substances". Some 30 years later, Hill³ turned this notion of an unspecified chemical substance into a specific mathematical operator, a quantity capable of interacting reversibly with an agonist. His mathematical description of drug activity laid the foundation for the quantitative analysis of competitive antagonism developed in the 1930s and 1940s.

The view that emerged led to the definition of a competitive antagonist as a compound which binds, reversibly, to a specific receptor. The process is governed by the law of mass action and is characterized by a single affinity parameter. Application of the idea is based on two assumptions: that agonist and antagonist compete for a single class of unoccupied receptors, and that the antagonist does not disturb the unknown relationship between receptor occupancy and tissue response so that this relationship can be cancelled out by using null methods. The inherent simplicity of the model has meant that deviations from a good fit are easy to detect.

To apply and test it, the physiological agonist is used to light up a specific class of target receptors so that not only can highly selective antagonist compounds be dis-

covered but their site of action can also be precisely specified. In this way, for almost 50 years, the model has proved to be invaluable for the development of new drugs, mainly competitive antagonists of physiological agonists (hormones, neurotransmitters, neuromodulators etc.).

Although a primitive concept of agonist activity is necessarily built into models of antagonism, explicit models of agonist action were not formulated until the 1950s and 1960s. They were developed to deal with the discovery of ligands, referred to as partial agonists, which were unable to produce maximum responses, giving rise to the notion of efficacy. The discovery of partial agonists led to the proposal that affinity and efficacy are separable quantities. There has been a renewed interest in agonist models to track and explain the variable expression of agonist activity among tissues and among species, which is important in new drug discovery.

Advances in biochemistry led to the isolation, characterization and sequencing of receptor proteins during the 1970s and 1980s, but turning receptor-as-a-mathematical-operator into receptor-as-a-chemical-entity did not undermine the classical models. In contrast, the more recent spectacular advances in molecular biology, leading to the isolation, expression and mutation of receptor genes, did so. Site-directed mutagenesis has been used to produce receptors which are active in the absence of agonist, and similar results have been achieved with wild-type receptors over-expressed in tissue culture.

Bond *et al.*¹ have now brilliantly developed transgenic mice in which a class of excitatory cardiac receptor is over-expressed by 200-fold. The animals' hearts behave as though they are being maximally stimulated even in the absence of the physiological agonist. It seems, then, that unoccupied receptors can exist in an active form. Some but not all of the ligands previously classified as simple competitive antagonists were unexpectedly able to switch off spontaneously active receptors, evidently a previously undisclosed property; this subclass of ligands are defined as inverse agonists.

At the heart of the traditional view of agonist activity is the idea that receptors have to be activated by occupation. In this sense, in the new model, agonists do not activate receptors (see Fig. 1 of the paper, page 272). Rather, their agonist activity is a consequence of binding to already active receptors, so that inactive receptors are pulled into activity by disturbing the equilibrium that exists between them. Inverse agonists bind to inactive receptors and pull the equilibrium the other way.

1. Pauly, D. & Christensen, V. *Nature* **374**, 255–257 (1995).

2. Vitousek, P. M., Ehrlich, P. R., Ehrlich, A. H. & Matson, P. A. *Bioscience* **36**, 368–373 (1986).

3. Alverson, D. L., Freeberg, M. H., Murawski, S. A. & Pope, J. G. *FAO Fish. Tech. Pap.* 339 (1994).

4. Kirkwood, G. P., Beddington, J. R. & Rossouw, J. A. in *Large Scale Ecology and Conservation Biology* (eds Edwards, P. J., May, R. M. & Webb, N. R.) 199–227 (Blackwell Scientific, Oxford, 1994).

5. FAO Fish. Circ. 853 (1993).

Competitive antagonists do not disturb the equilibrium because they have equal affinity for both receptor states. Thus, in this two-state model, the terms agonist and antagonist lose their traditional meaning and have to be replaced with ligands whose effect is governed by their differential binding for the two receptor states and by the number and ratio of these states in a particular tissue.

The consequence of the paper by Bond and colleagues is that pharmacologists can no longer ignore the implications of the two-state receptor model for the characterization and development of drugs. If the model is valid then there will be a need to develop over-expression systems for all

receptors to define the full range of properties and therapeutic potential of existing as well as new ligands. In so far as knowledge grows about disease states associated with constitutively active receptors, "inverse agonists . . . may represent an important and specific therapeutic approach for such disease states"¹. □

J. W. Black and N. P. Shankley are in the Department of Analytical Pharmacology, King's College School of Medicine and Dentistry, London SE5 9PJ, UK.

1. Bond, R. A. *et al.* *Nature* **374**, 272–276 (1995).
2. Langley, J. N. J. *Physiol.* **1**, 339–369 (1878).
3. Hill, A. V. J. *Physiol.* **39**, 361–373 (1909).

GALAXY MERGERS

Questions of clusters

Sidney van den Bergh

COLLISIONS (and subsequent mergers) between gas-rich galaxies result in violent bursts of star formation. Do such mergers simply increase the rate of normal star and cluster formation, or do they promote the creation of a special population of very massive clusters that will resemble the globular clusters in the halo of our Galaxy after about 10 billion years? New data that Whitmore and Schweizer¹ have obtained with the Hubble Space Telescope (HST) indicate the former, but the results are likely to be contentious.

Such questions have implications for galaxy evolution, as the processes associated with the formation of globular clusters are thought (or at least hoped) to shed light on the formation of galaxies. Under-

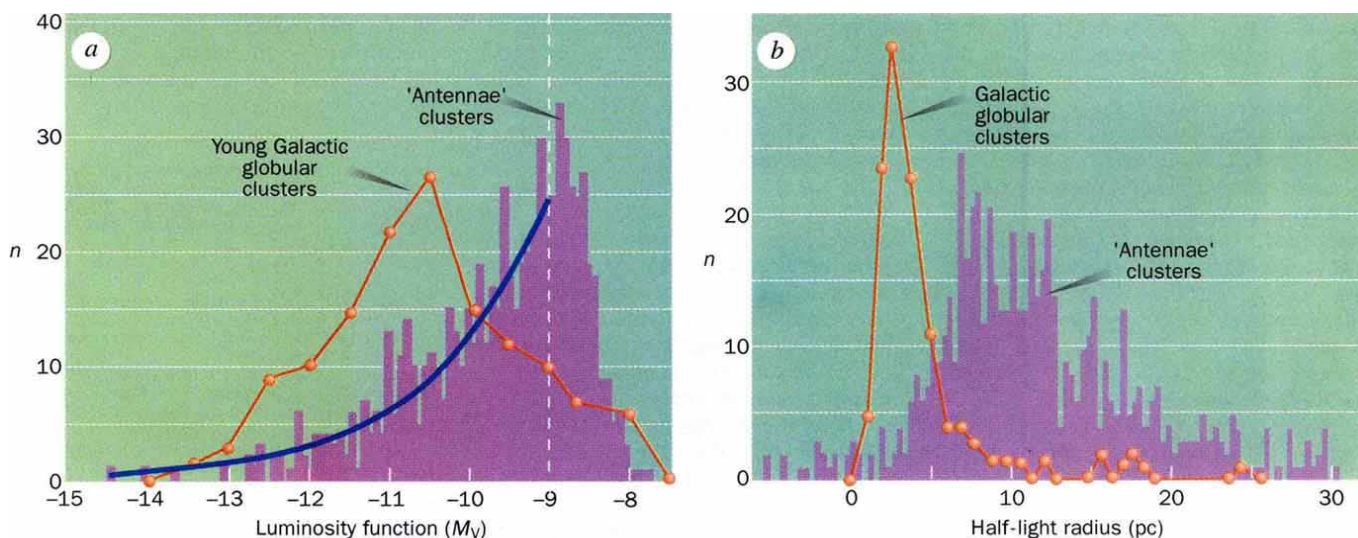
standing how gas can collapse rapidly to form such dense clusters of stars (several million within a radius of 10 parsec) might tell us what happened in the early Universe.

Whitmore and Schweizer's HST observations of the colliding giant spirals NGC4038 and NGC4039 ('the Antennae') have produced magnitudes, colours and radii for over 600 young clusters. The frequency distribution of luminosities for these clusters, assuming a Hubble constant² of 80 km s⁻¹ Mpc⁻¹ and a mean internal absorption (A_V) of 0.5 mag, is shown in part *a* of the figure. This shows that the young blue clusters in the Antennae have luminosities in the range $-14.5 < M_V < -8$.

Above the 'completeness limit' at $M_V \approx -9$ mag (the brightness below which faint clusters are lost in noise), the luminosity function of clusters appears to have a roughly exponential shape. In this respect, these clusters resemble open clusters^{3,4}, such as are found in the disk of our Galaxy, rather than globular clusters, which are observed to have a gaussian luminosity function in a wide range of environments.

The light of young clusters is dominated by massive blue stars, so unresolved young clusters will appear blue. As a result of stellar evolution the most massive stars will evolve and die first, and as these young stars disappear the cluster will become redder. On the basis of cluster colours, Whitmore and Schweizer find that the blue clusters in NGC4038 and NGC4039 typically have ages of around 40 million years. Using evolutionary calculations by Bruzual and Charlot⁵, one finds that globular clusters will fade by about 3.25 mag between ages of 4×10^7 and 1.5×10^{10} years. Applying this correction to the globular clusters in our Galaxy, one obtains the luminosity function shown by the line plot in part *a* of the figure, which clearly shows that the blue clusters in NGC4038/4039 do not have a globular-cluster-like luminosity function. This suggests that collisions between galaxies simply increase the rate at which normal stars and open clusters form in Galactic disks, and that they do not result in the preferential formation of a population of massive protoglobular clusters.

Of course, mass-loss by evaporation, and tidal mass-losses due to encounters with giant molecular clouds will, in time, preferentially destroy low-mass open clusters. But such effects cannot explain the



a, Comparison between the frequency distribution of luminosities for young blue clusters in the colliding galaxies NGC4038 and NGC4039 and the luminosity function that Galactic globular clusters would have had at a similar age of ~40 million years. These results suggest that the collision between these two galaxies has increased the rate at which open clusters are formed (that is, they imply that the

clusters are not young globular clusters). *b*, The half-light radii of clusters in NGC4038/4039 appear to be some three times larger than those of Galactic globular clusters. In each case, the histogram indicates the new results obtained by Whitmore and Schweizer¹, and the orange line shows the corresponding results for globular clusters in our Galaxy.

enormous difference between the masses of about 1×10^3 solar masses ($M_\odot$) of old open clusters such as M67 and NGC188 and those of typical globular clusters, which have masses of around $1 \times 10^5 M_\odot$.

Part *b* of the figure is a comparison between the half-light radii of clusters in NGC4038/4039 measured by Whitmore and Schweizer, and those of Galactic globular clusters. It shows that the blue clusters in the Antennae are typically three times larger than Galactic globular clusters. Similarly, Whitmore *et al.*⁶ find that the young blue galaxies in NGC7252 (the 'Atoms-for-Peace' galaxy, so called because of its shape) are larger than Galactic globular clusters. The differences are in the opposite sense to those expected from stellar evolution; this is because rapid mass-loss in fast-evolving young stars will cause the radii of young clusters to increase with time, thus increasing the difference between the radii of clusters in the Antennae and Galactic globular clusters. An important point, however, is that the observations of NGC4038/4039 and of NGC7252 were obtained before the HST was refurbished. These objects will have to be re-observed with the new, improved HST optics to confirm that young blue clusters in interacting galaxies are more diffuse than Galactic globular clusters.

The tentative conclusion that mergers form objects resembling open clusters rather than globular clusters bears upon currently fashionable ideas on galaxy evolution. In particular, it militates against the suggestion that most elliptical galaxies formed by mergers of spiral galaxies. It has been known for a number of years that the specific cluster frequency — that is, the number of globular clusters per unit luminosity — is much greater in ellipticals than it is in spirals. The most plausible way of accounting for this difference was to conjecture that most of the globular clusters observed in ellipticals were actually formed during collisions between spirals⁷ and were not primordial objects that formed at the same time as the ancestral spirals. The new observations by Whitmore and Schweizer provide *prima facie* evidence against this hypothesis; as such they will generate wide debate. □

Sidney van den Bergh is at the Dominion Astrophysical Observatory, 5071 West Saanich Road, Victoria, British Columbia V8X 4M6, Canada.

1. Whitmore, B. C. & Schweizer, F. *Astr. J.* **109**, 960–980 (1995).
2. Friedman, W. L. *et al. Nature* **371**, 757–762 (1994).
3. van den Bergh, S. & LaFontaine, A. *Astr. J.* **89**, 1822–1824 (1984).
4. van den Bergh, S. in *Structure and Dynamics of Globular Clusters* (eds Djorgovski, S. G. & Meylan, G.) 1–13 (ASP Conf. Ser. 50, Astr. Soc. Pacific, San Francisco, 1993).
5. Bruzual, G. & Charlot, S. *Astrophys. J.* **405**, 538–553 (1993).
6. Whitmore, B. C., Schweizer, F., Leitherer, C., Borne, K. & Robert, C. *Astr. J.* **106**, 1354–1370 (1993).
7. Schweizer, F. *Nearly Normal Galaxies* (ed. Faber, S.) 18–25 (Springer, Berlin, 1986).

New phase for mantle research

Dion L. Heinz

SILICON dioxide, or SiO_2 , in the form of quartz is the most ubiquitous mineral at the surface of the Earth. It has more than 15 polymorphs¹ that exist either at high temperature (seven exist at 1 bar), at high pressure or at simultaneously high temperature and pressure. Six years ago, the highest-pressure high-temperature polymorph known was the stishovite form, but in 1989, Tsuchida and Yagi² showed that at high enough pressures and temperatures, stishovite itself underwent a further structural transformation. On page 243 of this issue, Kingma *et al.*³ demonstrate that this transition actually occurs at lower pressures than originally thought. The phase transformation of stishovite to the phase with the 'CaCl₂' structure at 50 GPa could have profound repercussions for solid Earth geophysics: if its seismic signature can be found, at the corresponding depth of 1,200 km, this could answer the open question of whether free silica exists in the lower mantle.

Stishovite was made in the laboratory⁴ before it was found in nature (in Meteor Crater, Arizona)⁵. The unique feature of stishovite (which has the rutile structure) is that silicon occurs in octohedral coordination, unlike common silicates found at the surface of the Earth, where Si is tetrahedrally coordinated. Octohedral coordination of Si is the norm for lower-mantle silicates, as all upper-mantle silicate phases eventually transform to the perovskite structure in the lower mantle⁶ (see also ref. 7). In their 1989 experiments, Tsuchida and Yagi² found that the stishovite phase of silica transformed to the CaCl₂ structure at pressures above 108 GPa when the sample was laser-heated. They determined that the density change across the phase transition is about 1 per cent.

The motivation for the new experiments by Kingma *et al.* was the high degree of agreement between previously measured pressure dependencies of the Raman-active modes of stishovite (to 32 GPa)⁸ and linear augmented plane-wave (LAPW) calculations⁹. Furthermore, the LAPW calculations predicted that the transformation would occur at 50 GPa driven by the disappearance of the $C_{11}-C_{12}$ elastic modulus. As the change in volume is subtle, Kingma *et al.* used Raman spectroscopy to look for a characteristic change in the pressure dependence of the B_{1g} vibrational mode of stishovite predicted by theory, instead of X-ray diffraction techniques. Softening of this lattice vibration is related to a decrease in the $C_{11}-C_{12}$ elastic modulus.

Why the difference in the results this

time? The answer could lie in differences between the two experiments. The original experiments of Tsuchida and Yagi were non-hydrostatic and the samples had to be laser-heated to greater than 1,300 K for the CaCl₂ phase to be observed. Kingma *et al.* used a quasi-hydrostatic pressure medium (neon) and carried out all their experiments at room temperature. The non-hydrostatic nature of the original experiments could have caused an overestimate of the transition pressure. One possible reason that the CaCl₂ structure was only observed after laser-heating of the sample while at high pressures is that the process of laser-heating relaxed the pressure gradients in the sample.

The study by Kingma *et al.* suggests that free silica in the lower mantle would have the CaCl₂ structure at depths greater than 1,200 km. But does any free silica in fact exist there? The volume change from stishovite to the CaCl₂ structure would not be enough to stabilize an assemblage of mixed-oxide phases; rather, this would transform to a silicate perovskite assemblage^{2,7}. Several groups of researchers have concluded that the lower mantle is chemically different from the upper mantle^{10,11}, finding that the lower mantle is enriched in iron or silica, or both, relative to the upper mantle.

A simple but useful model is that the upper mantle's composition is $(\text{Mg,Fe})_2\text{SiO}_4$ with about 10 per cent iron, and the lower mantle is somewhere between $(\text{Mg,Fe})_2\text{SiO}_4$ (which would exist as silicate perovskite and magnesio-wüstite) and $(\text{Mg,Fe})\text{SiO}_3$ (which would exist as silicate perovskite), where added iron would increase the amount of Fe on the Mg,Fe site and added silica would move the lower-mantle composition towards

1. Deer, W. A., Howie, R. A. & Zussman, J. *Rock-Forming Minerals*, Vol. 4, 1–435 (Longman, London, 1963).
2. Tsuchida, Y. & Yagi, T. *Nature* **340**, 217–220 (1989).
3. Kingma, K. J., Cohen, R. E., Hemley, R. J. & Mao, H.-K. *Nature* **374**, 243–245 (1995).
4. Stishov, S. M. & Popova, S. V. *Geokhimiya* **10**, 837–839 (1961).
5. Chao, E. C. T., Fahey, J. J., Littler, J. & Milton, D. J. *J. geophys. Res.* **67**, 419–421 (1962).
6. Jeanloz, R. & Thompson, A. B. *Rev. Geophys. Space Phys.* **21**, 51–74 (1983).
7. Jeanloz, R. *Nature (News and Views)* **340**, 184 (1989).
8. Hemley, R. J. *Pressure Dependence of Raman Spectra of SiO₂ Polymorphs Alpha Quartz, Coesite, and Stishovite*, 1–486 (Terra Scientific/AGU, Tokyo, 1987).
9. Cohen, R. E. *First-Principles Predictions of Elasticity and Phase Transitions in High Pressure SiO₂ and Geophysical Implications*, 1–530 (Terra Scientific/AGU, Tokyo, 1992).
10. Stixrude, L., Hemley, R. J., Fei, Y. & Mao, H. K. *Science* **257**, 1099–1101 (1992).
11. Knittle, E., Jeanloz, R. & Smith, G. L. *Nature* **319**, 214–216 (1986).
12. Jeanloz, R. & Richter, F. M. *J. geophys. Res.* **84**, 5497–5504 (1979).
13. Knittle, E. & Jeanloz, R. *Geophys. Res. Lett.* **16**, 609–612 (1989).
14. Knittle, E. & Jeanloz, R. *Science* **251**, 1438–1443 (1991).
15. Goarant, F., Guyot, F., Peyronneau, J. & Poirier, J.-P. *J. geophys. Res.* **97**, 4477–4487 (1992).

(Mg,Fe)SiO₃. It does not seem that a silica enrichment alone would be enough to form free silica, but as the perovskite structure can only accommodate so much iron⁶ before it breaks down into an assemblage of magnesio-wüstite, silica and magnesium-iron-silicate perovskite, enrichment in iron and silica together could produce free silica.

If the transformation of silica from stishovite to the CaCl₂ structure can be observed seismically (it should be limited to a small depth interval, or appear seismically sharp, as it is a univariant phase transition), then the question of the composition of the lower mantle will be much closer to being answered. And if the chemical compositions of the upper and lower mantles turn out to be different, then convective flow between the upper and lower mantles will be inhibited¹². The form of convection in the Earth's mantle, either whole-mantle convection or layered-mantle convection, is a highly debated topic among geophysicists, and is intimately related to the chemical composition of the lower mantle.

The post-stishovite phase will definitely be important for the geophysical interpretation of the D'' region (the seismically

distinct layer at the bottom of the lower mantle just above the outer core). Recent experiments have shown that the reaction of molten iron and silicates at the temperatures and pressures to be found at the boundary between core and mantle produce a free silica phase¹³⁻¹⁵.

Perhaps the most exciting thing about the work by Kingma *et al.* is the strong interplay of theory and experiments. In this case the theory accurately predicted the pressure of the phase transition. This is wonderful to see, as stishovite and its high-pressure polymorph in the CaCl₂ structure are not highly symmetric simple structures. The LAPW calculations also predict that the CaCl₂-structured SiO₂ will undergo a further phase transformation to the pyrite structure above 150 GPa. This pressure range is well within modern capabilities, and thus experiments designed to detect this phase transition will provide a further check on the accuracy of the LAPW calculations. □

Dion L. Heinz is in the Department of the Geophysical Sciences and the James Franck Institute, University of Chicago, 5734 South Ellis Avenue, Chicago, Illinois 60637, USA.

DEVELOPMENTAL BIOLOGY

Growth factor lends a hand

J. M. W. Slack

In a remarkable series of experiments over 60 years ago, Balinsky was able to induce complete extra limbs by implantation of nasal tissue into the flank of amphibian embryos¹ (see figure). These results have since reposed in the shadowy world of half-forgotten things, both because they were inexplicable, and because they seemed unconnected to other work on limb development. Now a paper by Cohn *et al.*² in the latest issue of *Cell* has shown that extra limbs can be induced from the flank of the chick embryo by implantation of beads soaked in fibroblast growth factor (FGF). This work may represent a breakthrough in the understanding of the molecular basis of limb determination, and it will certainly focus attention on the early stages of limb development that occur before the appearance of a visible limb bud.

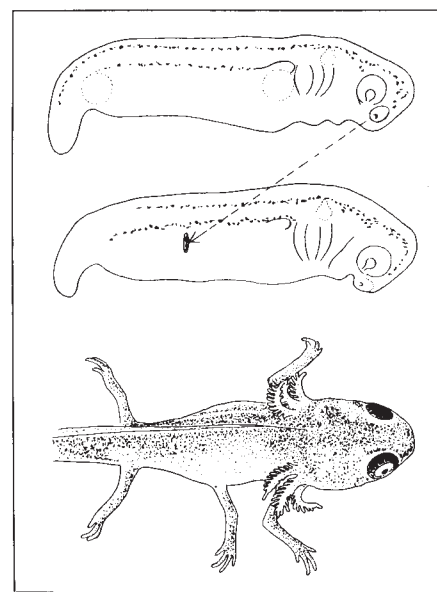
The basic result is that implantation of a heparin-acrylic bead, previously soaked in FGF-1, -2 or -4, into the flank between the prospective wing and leg levels, causes the formation of a new limb bud. The supernumerary buds are certainly real limb buds as they possess an apical ectodermal ridge (AER) and show the characteristic expression patterns of the genes *Sonic hedgehog* and *hoxd13*. After final differentiation, it can be seen that the

pattern of the skeleton is wing-like in those limbs induced near the normal wing, and leg-like in those induced near the normal leg. The girdles are present, as they were in Balinsky's induced limbs, indicating that development has commenced from the most proximal level. Most of the induced limbs have a reversed anteroposterior (thumb to little finger) polarity, something also seen in Balinsky's series. This is ascribed to the fact that the region capable of forming the wing's zone of polarizing activity (ZPA), the signalling centre that normally controls anteroposterior pattern, is on the anterior side of the new limb rather than the posterior side. The FGF induction works for a period of about eight hours, from about stage 13 when the head is beginning to turn, until stage 17 when the limb buds are becoming visible.

The outgrowth and patterning of the chick limb bud has been extensively studied in recent years and we are now approaching a molecular-level understanding of what is going on^{3,4}. But the FGF bead inductions force us to ask how the limb bud is formed in the first place. It is known that there are fore- and hindlimb rudiments within the lateral mesoderm, because limbs can develop from grafts of mesoderm from the future

bud positions implanted in other sites in the embryo⁵⁻⁷. But it is not really known when these limb rudiments come into existence or how they are formed. The school of Ross Harrison and his students, who worked with amphibian embryos, considered the limb rudiments to be established in the mesoderm as early as gastrulation⁸, but this may not be correct⁹. The problem is that if the limb arises as a result of interactions between several territories, then a large graft that includes all the territories is capable of forming a limb. But this does not mean that the limb rudiment itself existed before the graft was taken, merely that the conditions for its formation had been established within the graft area.

Indeed there is some evidence that a limb rudiment arises not too long before the bud begins to grow, and that interactions within the mesoderm are necessary for this to occur. In the chick, for example, mesoderm from the prospective wing level taken before stage 11 and grafted to the flank does not form a limb unless it is accompanied by some somitic tissue^{10,11}. This trophic requirement for somite tissue should not be confused with the well established contribution of somitic muscle cells to the limb bud, which occurs later and consists of a migration of myoblasts from somites to limb. In amphibian embryos, the mesodermal 'limb disc'⁸ consists of a small limb rudiment which needs to be accompanied by dorsal and posterior margin tissues for development on ectopic sites^{12,13}. These results all suggest that a mesodermal limb rudiment is formed during the hours before limb bud outgrowth, and that interactions are involved in its formation. ►



The results of Balinsky's experiments of 1933. An extra limb was induced following the implantation of a nasal placode into the flank of a newt embryo at the pre-limb-bud stage (reproduced from ref. 25).

The first visible step in limb bud development is the induction of the AER from the epidermis overlying the mesodermal limb rudiment. An AER can be induced from flank epidermis if a mesodermal limb rudiment is implanted beneath it⁵⁻⁷. The ability of prospective wing mesoderm to do this extends from about stage 11 to 17, from the stage of somite-dependence until the beginning of wing bud outgrowth. After this stage, the mesoderm of the bud loses AER-inducing activity, but acquires AER maintenance activity (now thought to be mediated by *Sonic hedgehog*), meaning that it can maintain a pre-existing AER. The competence of the flank epidermis to form an AER is lost by stage 19.

So how can we explain the formation of extra limbs from the flank following FGF treatment? Much attention has concentrated on the fact that the AER itself makes several kinds of FGF (FGF-2, -4 and -8) and it is probably this FGF that supports the outgrowth and patterning of the limb. But it is important to note that an AER cannot support limb development from flank tissue¹⁴, so the FGF bead effect must represent some earlier developmental event other than AER function. One explanation is that the FGF may replace the trophic influence from the somitic mesoderm, either as a direct signal or by enhancing the ability of flank mesoderm to respond to non-limb level somites. An alternative possibility is that FGF turns on ectopically some gene(s) responsible for defining subdivisions of the whole body pattern. This could mean that the territories that normally interact to cause limb formation are locally re-formed in the vicinity of the bead. It is already known that FGFs can activate some of the posterior Hox genes in *Xenopus*^{15,16}. An analogy to this 'whole body' explanation is a spectacular experiment showing that the mouse forelimb can be caused to become a mirror symmetrical duplication if the expression limit of *hoxb8* is brought forwards to the anterior side of the forelimb bud¹⁷. This happens because a ZPA is caused to form on both sides of the forelimb bud instead of just the posterior side, and so the bud is programmed to be posterior on both sides.

Having decided which interaction we think FGF is promoting, we then need to ask whether FGF itself is likely to be involved *in vivo* or whether it is mimicking some other factor. The less interesting possibility must be considered given that the doses used in the experiments are very high (1 mg ml⁻¹ applied to the bead) and might be regarded as pharmacological rather than physiological. A good start is to ask whether any of the known FGFs are expressed in a suitable position, such as the limb-level somites, or at a limb-level Hox boundary. There are very few data on the chick embryo itself, but it is known that FGF-2 (basic FGF) expression is widespread in the somite dermamyotome and in the flank epidermis at the appropriate stage, and is later concentrated in the AER¹⁸.

The other relevant expression data for the FGFs concern the mouse, in which limb buds with an AER first appear at the ninth day of development (E9). FGF-4 is found in the AER and in myotomes from about E9 (ref. 19). FGF-5 is expressed in the myotomes from E10.5 (ref. 20), and FGF-6 and FGF-7 from E9.5 (refs 21, 22). FGF-8 is found in the AER from E10.5 (refs 23, 24). So, as far as we know, there is no localized FGF expression in pre-limb-bud mesoderm. Expression of FGFs is frequently found in the myotomes, but in all cases apart from FGF-2 it is too late to be the trophic factor from the somites. The problem with FGF-2 is that its expression is very widespread, and it is still not clear whether it can be secreted from cells as it lacks a typical signal sequence. It may be that some as-yet unknown FGF is responsible, or that FGF can mimic the interaction causing the formation of limb rudiments, but does not actually cause it *in vivo*.

The recent study of FGF expression patterns has given us one other interesting piece of information: FGF-8 is now known to be strongly expressed in the nasal placode^{23,24}, so it may be that Balinsky himself was using FGF for his experiments back in 1933. □

J. M. W. Slack is in the ICRF Developmental Biology Unit, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

A broad view

THE traditional laser is a one-dimensional affair. Light passes back and forth along a cylindrical rod or cavity of lasing material, reflected by mirrors at each end. This apparently simple system can support many modes of oscillation, each with a slightly different wave pattern within the cavity. Selecting just one mode can be tricky.

Daedalus is now taking the multi-mode effect to extremes. His three-dimensional laser will be a sphere of lasing material completely surrounded by reflecting surface, a multilayer dielectric matched to the lasing frequency. This reflects the laser light very well, but allows the pumping radiation to enter the system unhindered. The volume laser will have an almost unlimited number of modes: back-and-forth diametral ones across any diameter, oblique ones bouncing round the sphere in any direction, whispering-gallery ones close to the surface and so on. Daedalus hopes to exploit this rich incipient chaos. His volume laser will be the ultimate telescope.

Imagine, he says, a big spherical laser on top of a mountain, exposed to the whole hemisphere of the night sky, and operating in pulsed mode. Consider the instant when the lasing cavity is fully pumped, its population of active sites highly inverted, and just about to lase. Any stray photon of the right frequency will trigger lasing. Each star in the sky will be sending a stream of photons through it from one precise direction. Those photons will be the 'seed' of a unique set of modes which, amplified millions of times by laser action, will encode precisely the direction of that star. The total ensemble of modes excited in the laser will be the sum of all these stellar contributions. The mode distribution of the resulting intense laser pulse will encode a perfect picture of the entire night sky.

To recover that picture, a retina of charge-coupled devices will study the bottom of the hemisphere, from which half the laser light will emerge, and relay their findings to the decoding computer. The volume-laser telescope will have the resolution of a lens of the same diameter, but (thanks to its super-regenerative amplification) almost ultimate sensitivity. Furthermore, it need not even be perfectly spherical. A more complex or imperfect shape could easily be allowed for in the decoding software. A volume-laser telescope a few metres across, taking in the whole sky at each blink of its laser eyeball, would soon present astronomers with their ultimate dream (or nightmare): a complete, definitive map of the sky.

David Jones

- Balinsky, B. I. *Wilhelm Roux Arch. EntwMech. Org.* **130**, 704-746 (1933).
- Cohn, M. J. *et al. Cell* **80**, 739-746 (1995).
- Niswander, L. *et al. Nature* **371**, 609-612 (1994).
- Lauffer, E. *et al. Cell* **79**, 993-1003 (1994).
- Kieny, M. J. *Embryol. exp. Morph.* **8**, 457-467 (1960).
- Dhouailly, D. & Kieny, M. *Dev Biol.* **28**, 162-175 (1972).
- Saunders, J. W. & Reuss, C. *Dev Biol.* **38**, 41-50 (1974).
- Swett, F. H. Q. *Rev. Biol.* **12**, 322-339 (1937).
- Stocum, D. L. & Fallon, J. F. *J. Embryol. exp. Morph.* **69**, 7-36 (1982).
- Pinot, M. J. *Embryol. exp. Morph.* **23**, 109-151 (1970).
- Kieny, M. *Ann. Embryol. Morphogenet.* **4**, 281-298 (1971).
- Slack, J. M. W. *Nature* **261**, 44-46 (1976).
- Slack, J. M. W. *J. Embryol. exp. Morph.* **57**, 203-217 (1980).
- Zwilling, E. *Cold Spring Harbor Symp. quant. Biol.* **21**, 349-354 (1956).

- Ruiz i Altaba, A. & Melton, D. A. *Nature* **341**, 33-38 (1989).
- Isaacs, H. V., Pownall, M. E. & Slack, J. M. W. *EMBO J.* **13**, 4469-4481 (1994).
- Charité, J. *et al. Cell* **78**, 589-601 (1994).
- Savage, M. P. *et al. Dev. Dyn.* **198**, 159-170 (1993).
- Niswander, L. & Martin, G. R. *Development* **114**, 755-768 (1992).
- Haub, O. & Goldfarb, M. *Development* **112**, 397-406 (1991).
- deLapeyrière, O. *et al. Development* **118**, 601-611 (1993).
- Mason, I. J., Fuller-Pace, F., Smith, R. & Dickson, C. *Mech. Dev.* **45**, 15-30 (1994).
- Heikinheimo, M. *et al. Mech. Dev.* **48**, 129-138 (1994).
- Crossley, P. H. & Martin, G. R. *Development* **121**, 439-451 (1995).
- Balinsky, B. I. *An Introduction to Embryology* 5th edn (Saunders, Philadelphia, 1975).

Amphibian breeding and climate

SIR — Although increases in greenhouse gases are considered certain to precipitate substantial climate change over the coming decades, there is almost no information on the effect of these events on wildlife. I have made some observations on the breeding cycles of amphibians that indicate changes consequent on steadily increasing winter and spring average temperatures. All of the amphibians native to Britain undergo spring migrations to breeding ponds to reproduce, and are thus typical of most temperate species. The three native anurans (*Rana temporaria*, *Bufo bufo* and *B. calamita*) demonstrate clines of spawning times, from the earliest in west or southwest England to the latest in the cooler north and east¹⁻³, that are thought to be adaptive and climate-related. Such amphibians might therefore be expected to reflect climate change in the timing of their breeding behaviour. Two sites in southern England have been studied intensively since 1978; the first, in Hampshire, has a population of natterjack toads *B. calamita*⁴. This species is at its northerly range limit in Britain, and might therefore be particularly responsive to climate change. The second, in Sussex, has populations of the frogs *R. temporaria* and *R. kl. esculenta*, the latter as an introduction also at its northerly range edge, as well as all three native urodeles *Triturus vulgaris*, *T. cristatus* and *T. helveticus*⁵.

The British anurans deposit spawn in conspicuous clumps that are easy to record. The two anurans at their northerly range edge spawned progressively earlier during the past 17 years (*a* in the figure). In both cases the trends were significant (for *B. calamita*, $r = -0.706$, $P < 0.01$; for *R. kl. esculenta*, $r = -0.583$, $P < 0.05$), and average times of first spawnings in the most recent 5 years (1990–94) were 2 and 3 weeks earlier, respectively, than in the first 5 years (1978–82). *R. temporaria*, however, showed no significant change over this period. The urodeles were assessed by first appearance of adults in the breeding pools. All three species showed dramatic trends towards earliness over the 17 years of observations (*b* in the figure), which were highly significant in all cases (for *T. vulgaris*, $r = -0.778$, $P < 0.001$; for *T. cristatus*, $r = -0.592$, $P < 0.02$; for *T. helveticus*, $r = -0.604$, $P < 0.02$). By 1990–94, the first newts were arriving in the ponds an average of 5–7 weeks earlier than in 1978–82.

These changing patterns of amphibian reproductive cycles correlated with changes in climate over the same period. To minimize the risk of type 2 statistical errors, correlations with climate data from the nearest recording stations were not

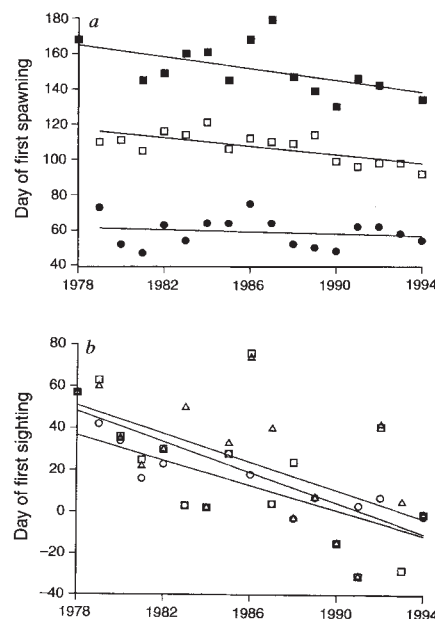
carried out randomly, but only to test specific hypotheses based on the physiology of amphibian reproductive cycles. Gametogenesis is temperature-dependent; in temperate amphibians it is completed either during the early spring after emergence from hibernation in the case of late breeders such as *B. calamita* or *R. esculenta*, or in winter for the remaining early-breeding species, and migration to breeding ponds requires threshold minimum temperatures and humidity levels.

I therefore looked for correlations between spawning or pond arrival time and average maximum or minimum temperatures, and rainfall, in the 1–2 months immediately preceding the observations and (in the case of early breeders) also from December to March. Striking correlations were found in all cases; thus spawning dates (lateness of spawning) of the April–June breeders *B. calamita* and

R. kl. esculenta were negatively correlated with average minimum temperatures in March + April ($r = -0.637$, $P < 0.01$) and maximum temperatures in March ($r = -0.769$, $P < 0.01$), respectively. Interestingly, the former species is predominantly nocturnal and the latter diurnal during activity periods in early spring. The spawning date of the early (February–March) breeding *R. temporaria*, which did not change significantly over the 17 years of observations, was nevertheless strongly negatively correlated with overall winter maximum temperatures over this period ($r = -0.796$, $P < 0.001$). Because the appearance of the three newt species in the ponds at site 2 were strongly inter-correlated, only the most abundant (*T. vulgaris*) was examined with respect to climate. As the changes of newt migration times were so dramatic, from first sightings in February at the start of the study to animals entering ponds in December by the end of it, it was not reasonable to make correlations with weather in the same month throughout the study as it was for the other species. For *T. vulgaris*, however, there was a strong negative correlation between lateness of pond arrival time and average maximum temperature in the month before arrival in any particular year ($r = -0.741$, $P < 0.001$). No significant correlations were found for any species with temperatures in other randomly selected months, and rainfall was also apparently unimportant.

The climate variables that correlate with amphibian breeding times themselves changed over the past 17 years much as might be anticipated from the data in the figure. Thus March and April average maximum temperatures at site 1 increased by an average of 0.11 °C per year ($r = 0.639$, $P < 0.01$), and at site 2 maximum temperatures in months prior to newt arrival by 0.24 °C per year ($r = 0.638$, $P < 0.01$). Winter average maximum temperature, the strong correlate with the *R. temporaria* spawn date that did not change much over time, also increased at site 2 by an average of 0.11 °C per year, but with much greater variance and thus as a trend that was only just statistically significant ($r = 0.542$, $P < 0.05$).

These observations suggest that amphibian reproductive cycles in temperate countries can respond sensitively to climate change. Thus spawning time of *R. kl. esculenta* and pond arrival time of *T. vulgaris* both became earlier during the course of these observations at rates of around 9–10 days per 1 °C increase in the



Temporal change in breeding activities. *a*, Days of year on which the first spawn was deposited by *B. calamita* at site 1 (□), and by *R. kl. esculenta* (■) and *R. temporaria* (●) at site 2. *b*, Days of year on which first individual *T. vulgaris* (○), *T. helveticus* (□) and *T. cristatus* (Δ) were observed in ponds in site 2. Ponds at site 1 in Hampshire, England, were examined every few days from the beginning of April every year after 1977 for the presence of *B. calamita* spawn as part of a long-term research programme on this species; date of first spawning was ascertained to within ± 2 days in each year. Ponds at site 2 in Sussex were examined every day and night from November to June (excepting periods of sub-zero temperatures), also every year post-1977, and days of first spawning by anurans and first appearances of urodeles were recorded with an accuracy of ± 1 day. Data were tested for normality and analysed by correlation.

1. Savage, R. M. *The Ecology and Life History of the Common Frog* (Pitman, London, 1961).
2. Cooke, A. S. *Br. J. Herpet.* **5**, 585–589 (1976).
3. Beebee, T. J. C. *J. Zool. (Lond.)* **205**, 1–8 (1985).
4. Banks, B., Beebee, T. J. C. & Denton, J. S. *Amphibia-Reptilia* **14**, 155–168 (1993).
5. Beebee, T. J. C. *Br. Herpet. Soc. Bull.* **17**, 12–17 (1986).

relevant maximum temperatures. Such changes in long-lived animals over such a short time must reflect behavioural and physiological plasticity rather than selection, and should be useful biological indicators of long-term climate change, as monitoring of the kind reported here is relatively simple to perform and surprisingly resilient to variation in population size.

Trevor J. C. Beebee

School of Biology,
University of Sussex,
Falmer, Brighton BN1 9QG, UK

Transgenic crops against parasites

SIR — The parasitic flowering plants *Orobanch* spp. (broomrapes) and *Striga* spp. (witchweeds) parasitize the roots of commercially important crops, heavily reducing yields. These parasites develop underground and emerge only when flowering. Although effective control methods are available for most weeds, there are only limited ways to control the root parasites¹, which exert their greatest damage before their emergence. The use of genetically engineered herbicide-resistant crops was tested as a solution to this intractable problem, as part of a project supported by the US Agency for International Develop-

ment Trilateral Egypt–USA–Israel Program. Here we show that the use of three target-site resistances prevented the growth of *Orobanch*, demonstrating a new potential use of transgenic herbicide-resistant crops.

Two kinds of resistance to herbicides can be genetically engineered into crops: metabolic resistance where the resistant plant degrades the herbicide to non-toxic products, and target-site resistance where the target enzyme affected by the herbicide is modified to preclude herbicide binding without changing the normal function of the enzyme. *Orobanch* and *Striga* spp. should be controlled early in their development, before causing crop damage. Target-site resistances permit safe movement of the herbicide from the host plant towards the underground parasites that are connected to its roots², whereas a crop with metabolic resistance would degrade the herbicide before it gets to the parasite through the host's conductive tissue. Metabolic crop resistances can therefore be useful only for herbicides directly applied to the parasite underground.

It has been suggested that the introduction of herbicide resistances to host plants would alleviate parasitism by treating infested fields with the herbicides^{2,3}. We tested four engineered resistances as models to examine our hypothesis: tomato plants engineered with the *bar* gene conferring metabolic resistance to glufosinate, an inhibitor of glutamine synthase⁴; tobacco plants with the *SuRB-Hra* gene encoding a modified acetolactate synthase (ALS) gene, conferring target-site resistance to chlorsulfuron⁵; oilseed rape (*Brassica napus*) plants containing the *aroA* gene encoding a modified enolphosphate-shikimate phosphate (EPSP) synthase, conferring target-site resistance to glyphosate⁶; and tobacco plants containing the *sul-I* gene which encodes a modified dihydropterotate synthase (DHPS) conferring target-site resistance to asulam⁷.

The application of glufosinate to the transgenic tomato plants (kindly provided by J. Leemans and P. Rüdelsheim of Plant Genetic Systems) failed to control *Orobanch*

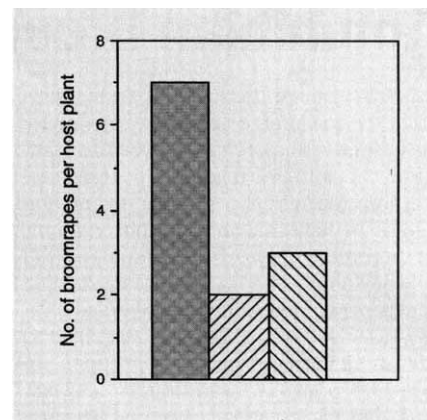


FIG. 2 Broomrape control by use of asulam-resistant transgenic tobacco. Foliar applications of 8 kg ha⁻¹ asulam at various times after planting. Left-hand column, non-treated; centre column, 4 weeks after planting; right-hand column, 6 and 8 weeks after planting.

aegyptiaca parasitizing its roots. Both the treated plants and the controls were heavily infected with the parasite. This was expected as the transgenic crop rapidly degrades glufosinate.

Conversely, the application of all three target-site resistances was effective in controlling the parasite. Broomrape was controlled on transgenic chlorsulfuron-resistant tobacco plants (kindly provided by B. Mazur and F. Lichtner of Dupont), by foliar application of chlorsulfuron 10, 20 or 30 days after planting in broomrape-infested soil. A single application of the herbicide resulted in 95% normal growth and flowering of the transgenic host (Fig. 1a). Similarly, glyphosate applied to glyphosate-resistant rape plants (kindly provided by G. Kishore and S. Padgett of Monsanto) infected with *Orobanch* completely prevented broomrape emergence, with concomitant normal growth and flowering of the transgenic crop (Fig. 1b). Broomrape was also controlled on transgenic tobacco plants with the target-site resistance to asulam (kindly provided by G. Freyssinet and B. Pelissier of Rhône-Poulenc).

Foliar application of asulam reduced the number of broomrape plants parasitizing a single host plant by 70% (Fig. 2). In all cases the control plants were heavily infected with the parasite, had less than half the normal height and shoot dry

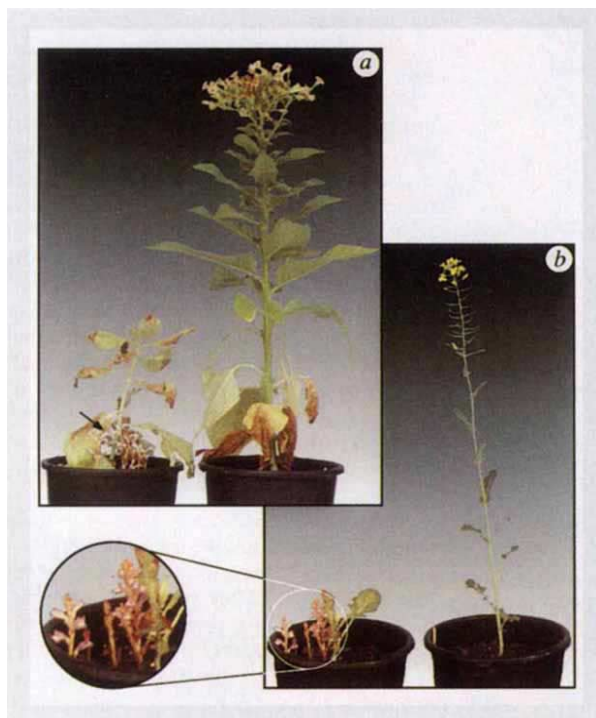


FIG. 1 Broomrape control by foliar application of herbicides on two transgenic herbicide-resistant crops grown in pots infested with about 5,000 seeds of *O. aegyptiaca* per kg soil. In each photograph the treated plant is on the right, the non-treated control with typical broomrape infection is on the left. *a*, Chlorsulfuron-resistant tobacco treated with 15 g ha⁻¹ chlorsulfuron; *b*, glyphosate-resistant rapeseed treated with 720 g ha⁻¹ glyphosate.

1. Parker, C. & Riches, C. *Parasitic Weeds of the World: Biology and Control* (CAB International, London, 1993).
2. Gressel, J. in *Resistance '91: Achievements and Developments in Combating Pesticide Resistance* (eds Denholm, I. et al.) 283–294 (Elsevier, London, 1992).
3. Foy, C.L., Jain, R. & Jacobsohn, R. *Rev. Weed Science* **4**, 123–152 (1989).
4. De Block, M. et al. *EMBO J.* **6**, 2513–2518 (1987).
5. Lee, K.Y. et al. *EMBO J.* **7**, 1241–1248 (1988).
6. DellaCioppa, G. et al. *Bio/Technology* **5**, 579–584 (1987).
7. Guérineau, F. et al. *Pl. molec. Biol.* **15**, 127–136 (1990).

weight, did not flower and died soon after the emergence of the broomrape flowering stalk.

The doubled yields afforded by control of the parasites will more than offset the added cost of both the transgenic seed and the small amount of herbicide, even in underdeveloped countries. This approach should only be used with crops that do not interbreed with related weeds in the same locality. The use of such transgenics represents a necessary stopgap measure until other means are found, as resistance can rapidly evolve to

some of these herbicides (such as the ALS inhibitors), but is expected to evolve more slowly to the others².

Daniel M. Joel

Yeshiahu Kleifeld

Dalia Losner-Goshen

Geza Herzlinger

Department of Weed Research,

Newe-Ya'ar Research Center,

Haifa 31900, Israel

Jonathan Gressel

Department of Plant Genetics,

Weizmann Institute of Science,

Rehovot 76100, Israel

Feather asymmetry in *Archaeopteryx*

SIR — After comparing flight feather asymmetry of *Archaeopteryx* with the asymmetry of several extant birds with various flight styles, 'flapping', 'gliding' and 'flightless', Speakman and Thomson¹ claim that *Archaeopteryx* was not capable of sustained flapping flight. In my view, their analysis is flawed and does not support their conclusion.

First, they report asymmetry values for *Archaeopteryx* feathers that are far too low, and they make a misleading comparison with extant birds. I have examined high-quality photographs in two fold-out plates², showing the left and right wing of the Berlin *Archaeopteryx* specimen at 2.8 and 4.1 times their natural size. The first three feathers are staggered in length, so the anterior margin is entirely free and clearly visible all along the outer half of feathers 1 and 2 in both wings. Also, feather 3 of the right wing has its anterior margin free in the outer part that extends beyond feather 2. The photographs show the ventral side of the wing, so, because of the way feathers are arranged in a bird wing, the rear margin of these feathers is also free and unobscured. At about 25% of the feather length from the tip, the feather shaft of the four anteriormost primary feathers (1–4) is located 24–34% of the feather chord behind the leading edge; thus, the trailing-vane is 3.11–1.91 times as wide as the leading-vane. This is much more asymmetrical than 41% of the chord behind the leading edge, or the ratio 1.46, reported by Speakman and Thomson for primary feathers 4, 5 and 6 of the Berlin specimen.

My asymmetry values are near the lower limit of, but partly inside, the range for 'flapping' and 'gliding' birds (Fig. 2 in ref. 1). Further, Speakman and Thomson's values for extant 'flapping' and 'gliding' birds are from primary feathers 1 or 2, which are more asymmetrical than feathers further back (Fig. 1 in ref. 1), but they nevertheless compare them with primaries 4, 5 and 6 in *Archaeopteryx*. I therefore conclude that the feather asymmetry of *Archaeopteryx* is within the range of modern birds using flapping flight.

Second, on the wing's downstroke, the

anteriormost primary feathers of modern birds often separate so that the outer part of each one acts as an aerofoil on its own³. The vane asymmetry then comes into operation and effects a proper orientation of the feather to the incident air stream. On random changes of angles of attack, the aerodynamic centre of pressure of a flat plate, or feather, does move fore and aft in such a way that it maintains dynamic stability in pitch, and hence in angle of attack, provided that the span-wise torsion axis lies within the interval 27–35% of the chord length behind the leading edge⁴. The feather shaft acts as a local torsion axis in any chord-wise profile. When the feather shaft lies ahead of the 27% chord point, there is a nose-down pitch moment that also tends to match the angle of attack to the incident air-stream.

The crucial feature of *Archaeopteryx* feather asymmetry is that the shaft of the anteriormost primary feathers is 24–34% of the chord behind the leading edge, completely ahead of the critical rear limit 35%, as required for self-stability in pitch. Consequently, vane asymmetry in *Archaeopteryx* primary feathers is pronounced enough to confer automatic pitch control on separated feather tips, and therefore does not indicate a lack of powered flight.

The asymmetry of the first three primaries is what matters most. Because of their staggered length in *Archaeopteryx*, they all form part of the wing's leading edge and therefore they are the feathers most likely to split apart in flight, acting as aerofoils — actually, leading-edge slats — on their own, with a need for the automatic pitch control that the vane asymmetry gives.

Third, the function of vane asymmetry is linked to the function of feather curvature⁵. Flight feathers in bird wings are per-

manently bent backwards and the feathers are free to rotate in the nose-up sense in their sockets. Nose-down rotation is prevented by the feathers' attachment in their sockets. It is also restricted by the nearest-neighbouring feather lying behind it, and partly overlapping the upper side of the trailing vane, so that the feathers are pressed together on the down-stroke. As a result of feather curvature, and less of vane asymmetry, the flight feathers of modern birds rotate in the nose-up sense on the upstroke when this is aerodynamically non-functional, letting air through the wing. This rotation occurs about an axis through the feather base, and because of the shaft curvature, most of the feather vane is behind this axis to provide the required nose-up torque.

Archaeopteryx flight feathers are as strongly curved as those of modern birds³. Its feather curvature alone could therefore provide the nose-up moment required for individual feathers to rotate about the axis through the feather base to let air through the wing on the upstroke. Also, the vane asymmetry in *Archaeopteryx* is pronounced enough to prevent a counteracting nose-down moment to be set up about the local feather shaft on the upstroke. Feather curvature and vane asymmetry in *Archaeopteryx* are therefore fully consistent with an active, flapping, flight mode.

R. Åke Norberg

Department of Zoology,

University of Gothenburg,

Medicinargatan 18,

S-413 90 Gothenburg, Sweden

SPEAKMAN AND THOMSON REPLY — Norberg raises several interesting but erroneous comments.

(1) He has measured vane asymmetry for one *Archaeopteryx*. His measurements exceed those made by us¹ on two specimens. Unfortunately, many of his measures were made on regions of the feathers which are overlapped. Since measuring asymmetry depends on defining both feather margins, his measurements depend critically on a subjectively inferred position of the hidden margin. Different observers have inferred asymmetry (a_p) for the overlapped feathers of *Archaeopteryx* ranging from complete symmetry⁵ ($a_p = 1.0$) to extreme asymmetry⁶ (a_p is approximately 4–5 from Fig. 1 in ref. 6). Our measurements are more accurate as they were made on sections of the feathers which do not overlap. The measures he made which were not overlapped refer to the diminutive first and second primaries; these small feathers would not be important in flight. In our measurements of extant birds we ignored the first primary when it was diminutive (less than half the length of the second primary); thus, comparing the diminutive primaries of *Archaeopteryx* to our sample of extant birds is invalid.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and five references.

If we make similar measurements on the photographs of a cast of the counter-slab of the Berlin specimen provided by Rietschel², which were used by Norberg, for regions of the fourth primary which are not overlapped, we obtain almost the same result as that we obtained previously¹ from measurements of a cast of the main slab (1.51 instead of 1.46). These measurements are also consistent across the only two specimens of *Archaeopteryx* where feathers are sufficiently preserved to make measurements. Although we previously measured several feathers, we included only measurements from feather 4 in the statistical comparison¹. We also showed that vane asymmetry at the 25% point does not vary between the first and fourth primaries of extant birds. Our measurements and statistical treatment are correct. However, if we accept Norberg's estimates of vane asymmetry, these clearly fall at the very lower end of the range for extant flying birds and well within the range of current flightless birds. Any inference that the bird was capable of powered flight from these data is extremely weak.

(2) The coincidence that Norberg's estimates of vane asymmetry fall within a hypothetical 'optimal' region is irrelevant to the discussion of flight performance. Many currently flightless birds have feathers with asymmetries which fall within this region: most extant birds using powered flight do not (Fig. 2 of ref. 1).

(3) Norberg provides no quantitative data concerning comparisons of vane curvature in current flying and flightless birds. However, the powered flight capabilities he infers from his vane asymmetry estimates and unsupported vane curvature are inconsistent with many other aspects of these fossils: in particular the wrist structure which is incompatible with execution of a flapping motion⁷, the pectoral muscles which were too small for powered flight⁸ and the long bony tail which may have been aerodynamically inefficient⁹. Rather than claim from our analysis that *Archaeopteryx* was not capable of sustained flapping flight, we stated that the poorly developed asymmetry was consistent with these other features.

J. R. Speakman

S. C. Thomson

Department of Zoology,
University of Aberdeen,
Aberdeen AB9 2TN, UK

Proton movement on membranes

SIR — Lateral proton diffusion along membrane surfaces represents certainly the most efficient system for proton movement between source and sink on a membrane. For the ATP-synthase activity in alkalophilic bacteria, this kind of diffusion may be a prerequisite because of the extremely low proton concentration in the medium. However, the proton diffusion along membrane surfaces has been a subject of continuous debate. Heberle *et al.*¹ used surface-bound pH indicators to measure light-induced pH changes on the extracellular and cytoplasmic surface of the purple membrane. The pH changes detected on the cytoplasmic side were slower than on the extracellular surface, the proton release side, but were clearly faster than those measured in the aqueous bulk phase.

These data were interpreted by the authors as the result of a fast proton migration along the purple membrane surface to the opposite side, with retarded surface-to-bulk transfer. Although this interpretation is certainly plausible, it is ambiguous and similar data could be obtained by a transient deprotonation on the cytoplasmic surface. Therefore, it is necessary to find direct experimental evidence for this preferential lateral proton migration. If the protons do move along the membrane surface, this lateral diffusion will most probably be mediated by the surface-bound buffer groups (acidic and basic surface residues, for example, lipid head groups). The dwell time of protons on lipid head groups (the inverse of the dissociation rate constant) was shown to increase rapidly with their pK_a (refs 2, 3). Therefore, the lateral proton movement along membranes would be expected to depend on the pK_a of the lipid head groups.

An experiment that would conclusively demonstrate lateral diffusion of protons along the membrane surface would show that this diffusion can be modulated by varying the chemical character of the lipid head groups. Detection of a proton, released on the extracellular side, should be affected exclusively on the cytoplasmic surface when lipids with different pK_a s are introduced. Unfortunately, such an experiment cannot be done easily with purple membrane without causing alterations in the photocycle and proton release kinetics. We characterized the proton movement in a bacteriorhodopsin-lipid-detergent micelle system⁴⁻⁶. The pH indicator fluorescein was covalently bound to cysteine residues, introduced into selected positions on the extracellular and cytoplasmic surface of bacteriorhodopsin by site-directed mutagenesis through reaction with the iodoacetamido derivative of fluorescein^{4,5}. In this system, the light-induced proton concentration

changes detected on the extracellular and cytoplasmic surfaces were also clearly faster than in the aqueous bulk medium. Moreover, this micellar system allowed us to exchange lipids without affecting the photocycle kinetics.

The proton mobility along the micellar surface could be varied, as predicted, by adding phospholipids with head groups of different pK_a s (refs 4, 6). Head groups with a low pK_a enhanced the lateral movement of protons from the extracellular release side to the cytoplasmic surface, resulting in a larger amplitude for the proton signal measured on the latter side. When lipids with higher pK_a head groups were introduced, the amplitude of the proton signal was considerably reduced. No changes were observed on the extracellular side. These results clearly demonstrate that the proton released on the extracellular side does move along the micellar surface to the cytoplasmic side and the role of the pK_a of lipid and detergent head groups. The protons move significantly faster from the extracellular to the cytoplasmic side than from the micellar surface to the aqueous bulk phase. Furthermore, we could demonstrate a rate-limiting step for the lateral proton movement from the protein to the lipid/detergent surface⁷.

We have also investigated the proton movement in the purple membrane, applying the technique described above for the micellar system using mutant purple membrane with cysteine in selected positions⁸. In principle, our results do agree with the interpretation presented by Heberle *et al.*¹, and the proton is also detected faster on either surface of the purple membrane than in the aqueous bulk phase. In our case⁸, however, the time difference for detection of the released proton between the two surfaces is smaller, with $71 \pm 4 \mu s$ on the extracellular and $76 \pm 5 \mu s$ on the cytoplasmic side compared with 76 ± 18 and $228 \pm 39 \mu s$, respectively, reported by Heberle *et al.* Based on our data, the surface diffusion constant is much larger than estimated by Heberle *et al.* ($D > 3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ compared with $D = 9.6 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$).

Peter Scherrer

Department of Biochemistry,
The University of British Columbia,
Vancouver, British Columbia V6T 1Z3,
Canada

1. Heberle, J., Riesle, J., Thiedemann, G., Oesterheld, D. & Dencher, N.A. *Nature* **370**, 379–382 (1994).
2. Nachliel, E. & Gutman, M. *J. Am. chem. Soc.* **110**, 2629–2635 (1988).
3. Gutman, M. & Nachliel, E. *Biochemistry* **24**, 2941–2946 (1985).
4. Scherrer, P. *et al. Struct. Funct. retinal Prot.* **221**, 205–211 (1992).
5. Alexiev, U., Marti, T., Heyn, M.P., Khorana, H.G. & Scherrer, P. *Biochemistry* **33**, 298–306 (1994).
6. Scherrer, P., Alexiev, U., Marti, T., Khorana, H.G. & Heyn, M.P. *Biochemistry* **33**, 13684–13692 (1994).
7. Alexiev, U., Marti, T., Heyn, M.P., Khorana, H.G. & Scherrer, P. *Biochemistry* **33**, 13693–13699 (1994).
8. Alexiev, U., Scherrer, P., Moliaaghbabab, R., Khorana, H.G. & Heyn, M.P. *Proc. natn. Acad. Sci. U.S.A.* **92**, 372–376 (1995).

1. Speakman, J. R. & Thomson, S. C. *Nature* **370**, 514 (1994).
2. Rietschel, S. in *The Beginnings of Birds* (eds Hecht, H. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P.) 251–260 (Freunde des Jura-Museums, Eichstätt, 1985).
3. Norberg, R. Å. in *The Beginnings of Birds* (eds Hecht, H. K. *et al.*) 303–318 (Freunde des Jura-Museums, Eichstätt, 1985).
4. Norberg, R. Å. *Biol. Rev.* **48**, 561–596 (1973).
5. Lowe, P.R. *Ibis* **86**, 517–543 (1944).
6. Fedducia, A. & Tordoff, M. *Science* **203**, 1021–1022 (1979).
7. Vasquez, R.J. *J. Morph.* **211**, 259–268 (1992).
8. Speakman, J.R. *Evolution* **47**, 336–340 (1993).
9. Baimford, A. *et al. Nature* **361**, 628–631 (1993).

Tackling Torquemada

Christopher Edens

Deciphering the Indus Script. By Asko Parpola. Cambridge University Press: 1994. Pp. 374. £60, \$95.

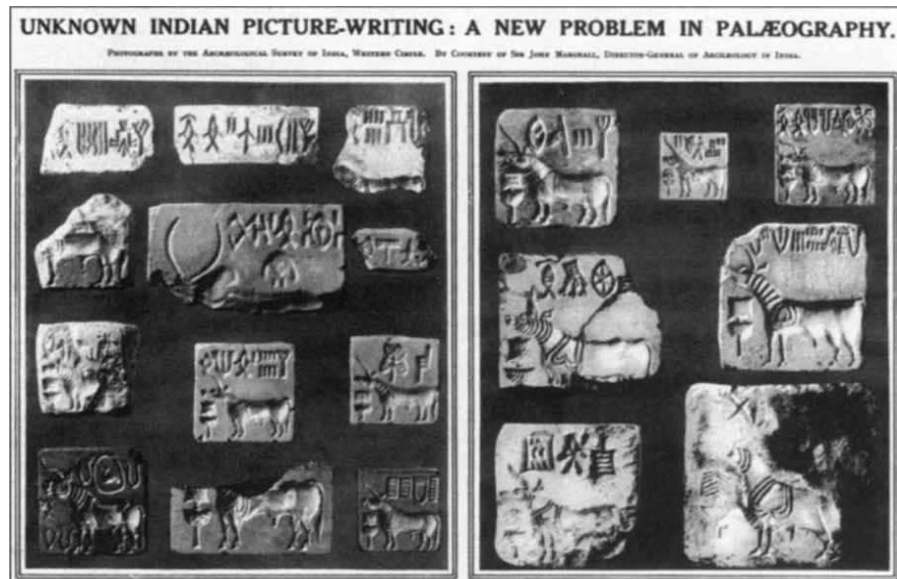
In 1924, following the first exploration of Harappa and Mohenjo-Daro in what is now Pakistan, John Marshall announced the discovery of the Harappan civilization in the *Illustrated London News*. Subsequent exploration revealed settlements of this civilization not only in the Indus valley but also into Gujarat in western India and eastward to Delhi. It now seems that the Harappan, or Indus, civilization came together suddenly around 2,500 BC and lasted about five hundred years. Among its most striking characteristics is a writing system that has been preserved as short inscriptions on seals and other artefacts. The script is unlike any Indian writing of later times or the mostly cuneiform scripts used in literate Bronze Age cultures of the Near East. Attempts to read the Harappan script have so far met with little success.

Asko Parpola belongs to a Finnish team that has been working on the Indus script since the 1960s. The computer-aided Finnish studies have produced several basic research references such as the *Corpus of Indus Seals and Inscriptions* and the *Concordance to the Indus Inscriptions*. The problem, according to Parpola, is that the successful decipherer of an unknown script must know either the language it represents or the values of its individual signs — but neither is known for the Indus script. After addressing several basic issues — how many signs there are in the script (a sign list), how to recognize compound signs, what direction to read inscriptions — Parpola stakes out three premises that earlier workers have also endorsed. First, he argues not only that the Indus language belongs to the Dravidian language group (and so is related to Tamil) but also that the Harappan culture had strong continuities with Vedic and Aryan India. Second, he uses sign counts to argue that the Indus script is 'logo-syllabic' in character, so that signs may stand for entire words and for syllables. This common kind of writing uses the rebus principle, so that homophones are represented by the same sign (for example the picture of an eye meaning 'eye', 'aye' and 'I' in English). And third, he suggests that the inscriptions consist mostly of the names of gods or theophoric personal names, as well as qualifying titles or attributes. This suggestion stems from analogy with Mesopotamian seals, the assumed function of the Indus seals and the brevity of the inscriptions.

Parpola then weaves together clues

from linguistics, religion, comparative folklore and archaeology to read individual signs. He begins by considering the common fish-shaped sign. 'Fish' is *m̃n* in most Dravidian languages, whereas the Proto-Dravidian root **m̃n* also means 'star'. Parpola argues that the Indus religion thought of its gods as stars, and wrote their names with the fish sign; more specifically, the plain fish sign denoted a prototype of Siva, with whom Hindu texts

stretches to propose a reading. For example, he begins his interpretation of the compound sign 'fish with internal dot' by finding Indian meanings — fertility, victory, sacrifice, destruction and illness — in the iconography of the 'fig deity' seal. This analysis leads him to expect a divine name that corresponds to Durga, the Hindu goddess of victory. Believing that the sign 'fish with internal dot' represents this name, he then searches for the star most likely to represent the proto-Durga. Parpola identifies the dot in the compound sign as *pot̃u*, 'red dot', which has the Central Dravidian homophone **pot̃u*, a kind of fish that in some languages corresponds to the Sanskrit *rohita*, the red carp. *Pot̃u* is also the Tamil word for the red mark painted on the foreheads of married women as a sign of sexual maturity and



Written in stone — pictures from the *Illustrated London News* (20 September 1924) of "pictographic seals" from Harappa and Mohenjo-Daro. They accompanied Sir John Marshall's article announcing "new discoveries of an unknown prehistoric past in India".

associate fish. Extending this basic finding, Parpola then identifies the common sign cluster 'six short strokes + fish' with the Old Tamil *āṇu-m̃n* or 'the six (visible) stars' of the Pleiades, and the sign cluster 'seven short strokes + fish' with *ēlu-m̃n*, the 'seven stars' of Ursa Major. A fish with a line across its body represents halving, the Proto-Dravidian root **pacu*, 'to halve', a homophone of the Proto-Dravidian **pacu*, meaning 'green'; consequently, the sign 'halved fish' is read **pacu-m̃n*, which relates to the Old Tamil *paccai* and refers both to the planet Mercury and to the Hindu god Krishna-Visnu. And a 'caret' (^) over the fish represents a roof, the Proto-Dravidian root **vey* or **mey*, 'to roof', which in turn has **may*, 'black' as a partial homophone: the roofed fish sign can be given the meaning **may-m̃n*, or 'black star', an Old Tamil name for the planet Saturn.

When Dravidian homophones are not readily available, Parpola sometimes

fertility, a reflection of the seal's iconography. The *rohita* fish, on the other hand, is associated in some Sanskrit texts with the cult of the Goddess, whose astral form is Rohiṇī, the star Aldebaran.

In these various ways, Parpola attempts to read a total of 24 individual and compound signs, which represent the names of astral deities manifested in various constellations, stars and planets, as well as the titles 'servant' and 'overseer' used by their human attendants. Parpola's book is very much a work in progress, and his proposed 24 readings seem a meagre return on the 278 pages of text. But the virtuosity of his arguments compensates for this imbalance, and Parpola leads the reader along a very pleasant, if sometimes slippery, path through the tangled details of the Harappan civilization, the linguistic history of India, and the myths, literature and rituals of Tamil, Vedic, Hindu and Buddhist tradition. Parpola is not alone in supposing strong continuities between the

Indus civilization and historical India, but his case sometimes strains credulity — as in his reading of the ‘fig deity’ seal. Another is his attribution of the Vedic nakshatra calendar to the Indus culture, which Parpola contends was compiled when the Pleiades rose heliacally at the vernal equinox during the late third millennium BC; he suggests that an earlier version of the calendar belongs to the late fourth millennium BC when Aldebaran (Rohini) rose at the vernal equinox. Bombarded with the hail of facts that make up these arguments, the reader is sometimes reminded more of *The Golden Bough* and enthusiastic Victorian scholarship than of the stodgy late-twentieth-century academy. Parpola compares decipherment to doing a crossword puzzle, in which “the probability of correctness increases with the number of interlocking solutions”. This book has lightly pencilled in the solutions to a few lines in one corner of a very large puzzle set by Torquemada. □

Christopher Edens is at the Peabody Museum of Archeology, Harvard University, Cambridge, Massachusetts 02138, USA.

Privileged order

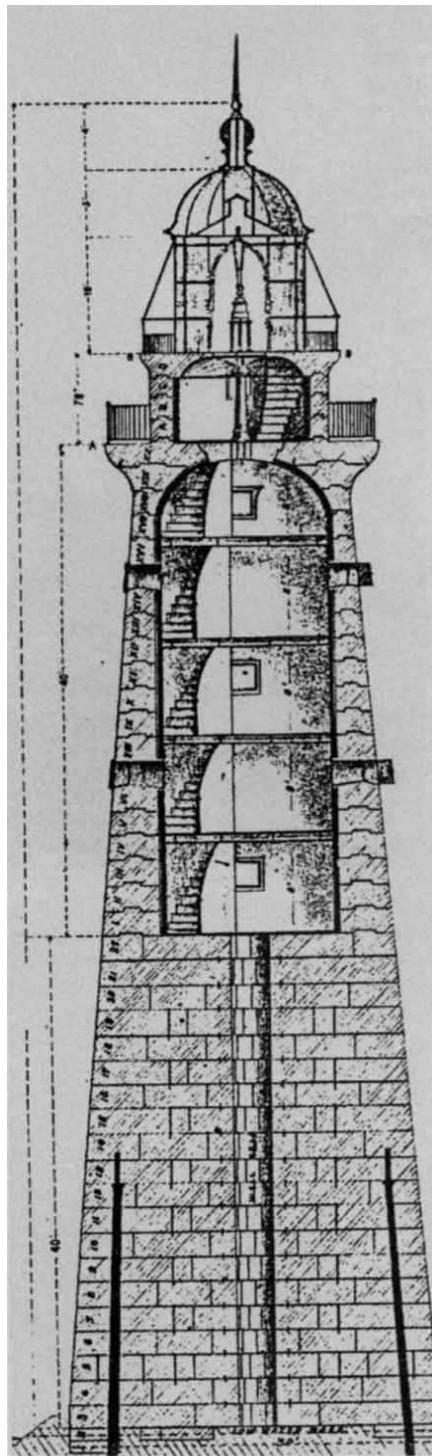
Henry Petroski

Structures in the Stream: Water, Science, and the Rise of the US Army Corps of Engineers. By Todd Shallat. University of Texas Press: 1994. Pp. 276. \$34.95.

UNTIL the later part of the eighteenth century, when John Smeaton declared himself to be an independent professional — a “civil engineer” — virtually all organized engineering was effectively rooted in military institutions. Indeed, earlier in the century, in 1716, the Corps des Ponts et Chaussées had been founded to take responsibility for civil engineering projects undertaken by the French government, and by the middle of the century the Ecole des Ponts et Chaussées had been created to train recruits to the corps. Formal engineering education was therefore established in France a full seven decades before it existed in Smeaton’s England, where the apprentice system prevailed well into the nineteenth century.

So two distinctly different cultures of modern engineering had been established in Britain and France by the time a young America was forming its own institutions. This background is nicely presented by Todd Shallat in his engaging history of the development of the US Army Corps of Engineers, which was strongly influenced by the scientific French rather than the experience-based British approach to engineering education and practice. For

example, in the early years of the US Military Academy, established at West Point in 1802, French textbooks predominated and so a necessary prerequisite of studying construction was the study of the French language. And the French influence went well beyond the academic, affecting even the nature of the uniforms adopted at West Point.



Faulty tower? Cross-section of the second Minots Ledge Lighthouse, near Boston, Massachusetts. The tower was completed by the US Army Corps in 1860. The first one collapsed in 1851 in high winds and seas a year after being built.

From Structures in the Stream

The Military Academy developed into a training ground for topographical engineers, who played such an important role in surveying and mapping the waterways and lands of the young country, and Shallat’s book reproduces many of the eighteenth-century portraits of the uniformed military engineers who were responsible for them. The maps tend to give the reader little more than an impressionistic flavour of the fruits of the endeavour, however, and the portraits give little insight into the personality of the engineers. (The format of the book’s illustrations tends to be small and some of the maps are so faint as to be almost unreadable.) Many a reader may wish that Shallat had used more words instead of these pictures, for his writing style is graphic and effective, and he has a considerable talent for placing what otherwise might seem to be the most mundane of engineering projects into cultural, technical and political contexts that provide much insight into the conflicts that shaped the Corps of Engineers.

Shallat’s story of the rise of the corps effectively ends with its conflict with James B. Eads, whose scheme of jetties for maintaining a shipping channel at the mouth of the Mississippi below New Orleans was bitterly opposed by the corps’ leaders, especially Andrew A. Humphreys, whose report with Henry L. Abbot on the physics and hydraulics of the Mississippi Eads had panned in a review. There was no love lost between Eads and the corps, which had also opposed his bridge across the Mississippi at St Louis, Missouri, to the point of recommending that it be torn down or that a canal be constructed around it. Such unreasonable and wrong-headed stubbornness on the part of the corps is what some critics of the institution today still find in the “privileged order of the very worst class”, in the words that Alden Partridge, a former West Point superintendent, used to characterize the corps in 1830. Shallat uses this condemning quotation as the title of his final chapter, which he concludes in the same more or less ambivalent tone with which he opens his book with regard to judging the much-judged corps.

Shallat has written a likeable and eminently readable book in which he lays out generally fairly and objectively the conflicts that he sees embedded in the origins and traditions of the Corps of Engineers and in the political structure in which it is embedded. But in the 200-odd pages of text, he (or his peer reviewers, his editor or his publisher) has not given himself enough space to tell the story in as much detail as it deserves. □

Henry Petroski is in the Department of Civil and Environmental Engineering, Duke University, Box 90287, Durham, North Carolina 22708-0287, USA.

Out of the flames of phlogiston

Bernadette Bensaude-Vincent

In the Shadow of Lavoisier: The *Annales de Chimie* and the Establishment of a New Science. By Maurice Crosland. *British Society for the History of Science*, 31 High Street, Stanford in the Vale, Farringdon, Oxfordshire SN7 8LH, UK: 1994. Pp. 333. £9, \$17 (pbk).

THE *Annales de Chimie*, a periodical founded in France in the revolutionary year of 1789, pioneered the modern style of scientific specialized journalism and remained one of the most important publications in chemical research throughout the nineteenth century. Although the title was inspired by a German model, Crell's *Chemische Annalen*, which started in 1778, the origin of the *Annales de Chimie* was closely linked to the overthrow of the phlogiston theory shaped by the Prussian chemist Ernst Georg Stahl — a revolution sparked by Lavoisier. At the time, the favourite journal for the rapid publication of research results was the *Observations sur la Physique*, and its editor, Claude de la Méthérie, was one of Lavoisier's fieriest adversaries. Supporters of the new chemistry, who had originally planned to publish a French translation of Crell's *Annalen*, used the new journal as an instrument for the propagation of Lavoisier's reformed chemical nomenclature and oxygen theory. Yet, although the *Annales* emerged from the flames of a controversy, from the very beginning it was much more than a tactical weapon for a new research school.

Maurice Crosland emphasizes the crucial influence of *Annales* in the making of modern chemistry. The first series of the journal, which was devoted to pure and applied chemistry, helped to shape the image of chemistry as a discipline related to physics although still closely associated with pharmacy. Because the editorial board, formed by the leading chemists of the time, encouraged readers to contribute to the *Annales*, the periodical proved crucial in raising the consciousness of a chemical community.

The title of Crosland's book is unfortunate in that it reinforces the traditional cliché of Lavoisier as the founder of modern chemistry. The contents of the book, however, suggest a much more complex view of nineteenth-century French chemistry. Whatever the prominent influence of Lavoisier's close collaborators — Fourcroy, Guyton de Morveau and Berthollet — on the first series, it is one of the great merits of Crosland's detailed historical study that it emphasizes the differences of scientific and ideological views among

these leaders and shows the deep changes in editorial policy introduced by the second generation of editors, Gay-Lussac and Arago, who coupled chemistry and physics in the title of the *Annales* in 1816.

Despite his attempt in the last chapter to discuss the enduring authority of Lavoisier in nineteenth-century French chemistry, Crosland has done more than most other scholars to disentangle the disciplinary dynamics of French chemistry from the heroic figure of Lavoisier. In the 1960s, under the influence of Henry Guerlac at Cornell University, the historiography of the chemical revolution tended to focus on Lavoisier's works. Crosland, however, worked on the fringe, investigating anonymous and obscure collective achievements such as the projects for reform of the chemical nomenclature (*Historical Studies in the Language of Chemistry*, 1962) or the informal research school founded by Laplace and Berthollet (*The Society of Arcueil*, 1967). More recently, he has devoted a study to the institution that dominated French science in the eighteenth and the nineteenth centuries (*Science under Control: The French Academy of Sciences*, 1992; for a review see *Nature* 357, 652 (1992)). Crosland has in this way gradually reconstructed the network of cognitive and institutional structures that determined French natural sciences in the early nineteenth century and has stayed clear of the historiographical debates about 'continuity versus discontinuity' or 'internal versus external history' that have trapped so many scholars over the past decades.

Regardless of his distance from traditional intellectual history, however, Crosland cannot be considered to be a pioneer of social studies of science. Indifferent to the current issues of the social construction of knowledge, he earnestly collects information from a variety of sources. Neither in this volume nor in his previous writings does he try to discover a narrative style. And the beautiful computer-drawn charts and diagrams that fill the pages of young scholars' dissertations hold no fascination for him. His main concern is not only to provide detailed information about the *Annales*, based on the manuscript records of the editorial board, but also to describe the context in which the *Annales* flourished and declined, with a thorough survey of various scientific journals published in nineteenth-century Europe.

With his old-fashioned Baconian style, with its abundance of details sometimes threatening the focus and unity of the volume, Crosland's study is an indispensable reference for the historiography of French science. Among the general issues for which it might be extremely useful is the disciplinary history of chemistry, especially its relationship with physics. Also, in contrasting the *Annales de Chimie*, sup-

ported by the establishment, with the rival *Journal de Physique*, which came to represent all outsiders or scientists on the fringe, and in stressing the lasting connection between the editorial board of the *Annales* and the French Academy of Sciences, Crosland sheds light on the still obscure phenomenon of the scientific orthodoxy that characterized French science in its heyday. And, finally, thanks to the information on the circulation of the *Annales* (600 copies in the first series), its business success (FF3,000 profit per editor in the second series) and above all the careful survey of a number of scientific journals in Europe, the book is a major contribution to the history of scientific journalism.

Bernadette Bensaude-Vincent is in the Département de Philosophie, Université de Paris X, 200 Avenue de la République, 92001 Nanterre Cedex, France.

The uses of beauty

A. D. Van Nostrand

Confessions of a Technophile. By Lewis M. Branscomb. *American Institute of Physics Press*: 1995. Pp. 255. \$29.95, £22.50. Distributed in the United Kingdom by Oxford University Press.

ALTHOUGH scarcely penitential, the confessions in this collection of essays are clearly declarations of belief and commitment. Lewis Branscomb believes that research scientists have a deep obligation to give something back to humanity in addition to the new knowledge that their work generates. His book proceeds from this premise, dealing with how he has experienced this obligation and with his "views on how society and scientists, politicians and experts, must work together to bring progress for humanity".

His views have evolved during a long and resonant career. A research physicist for 20 years, Branscomb later became director of the US National Bureau of Standards, editor of the *Reviews of Modern Physics*, chief scientist at IBM and a professor at Harvard's John F. Kennedy School of Government. Three US presidents have appointed him to science advisory posts and, drawing on the experiences of these appointments, he has co-authored two books and edited another on the subject of national technology policy. But the essays in the present volume are only incidentally autobiographical. Essentially they bring to bear on the subject of public policy Branscomb's insights into the mutual dependency of science, technology and society.

Twenty-three essays comprise the book,

eighteen of them published between 1968 and 1993 and somewhat revised; the others appear for the first time here. They deal with the way in which science works and with its social, political and economic consequences. Arranged in four clusters, they gradually develop an agenda for discussing public policy. The first cluster, "Unlikely Beginnings", addresses Branscomb's 20 years of research in atomic and molecular negative ions. "Thoughts about Science", the second cluster, deals with relationships between science and certain value-laden concepts, including integrity, complexity, aesthetics and societal expectations. The third cluster, "Motors for the Mind", concerns the development of computers from the past hundred years to a hundred years hence, with emphasis on the interactivity of these information machines and their users. "Taming Technology", the fourth cluster, raises the question of what we want our technology to do for us, and it frames a range of answers in terms of national debate and public policy.

An early item on Branscomb's agenda is the distinction commonly made between basic and applied research. He acknowledges its convenience for channelling resources into both scientific inquiry and the solution of social problems. But when based on some presumed motive of the investigator, he argues, the distinction is grossly misleading. It conveys different standards of intellectual excellence and, worse, it confuses the notion of risk: the term 'applied' conveys the idea that the practical benefits of inquiry are predictable and that they can be purchased with minimum risk.

Branscomb proposes a different logic. All good science is useful, he points out, and all research entails the asking of new questions. So in the sense that 'applied' means 'used', 'applied research' is an oxymoron. Contemplating this notion, he says, "conjures up in my mind a picture of Albert Einstein, hair flying, and underneath, the caption: 'Would you buy a used theory from this man?'". Branscomb is arguing for a better understanding of the holism of science and technology. Their continuous interactivity, he insists, is a reality that must be taken account of in framing the options of public policy, a reality that universities and other research sponsors have an obligation to instil in young investigators.

The confessions of this technophile are disarming. Not surprisingly, Branscomb has an acute sense of evidence, and with an extraordinary range of literary and historical references his essays convey an unmistakable authenticity. Yet the pieces themselves are so informal, so comfortable, as to convey a conversational tone. The collected conversations are also discursive, which is a limitation of the genre if one is looking for a straight-arrow argument. But there is more here than at first meets the eye; the subtle line of reasoning gradually conveys the mutual dependencies that exist between science, technology and society. Branscomb means to explain how society can use both science and technology for its survival. He sees an urgent need for society's policy-makers to understand how science and technology work: how they reinforce each other and how they differently serve human needs.

In the epilogue, "Science, Technology and Humanity", Branscomb makes his argument explicit by discussing several basic predications: that usefulness and beauty, the two most conspicuous attributes of scientific work, are not conflicting values, despite what many people believe; that the usefulness of technology is more immediate and tangible and the usefulness of science more conceptual; and that science outlines the limits of the technological future, providing us with a sense of diversity among options and with insights for evaluating technological alternatives.

Branscomb focuses this epilogue on the US political scene, but his remarks are applicable to most societies today. Science raises long-range questions, he observes, but political rewards lie in the short-haul. And there's the rub. To ask the right questions about technology, both science and politics share a need for feedback, for pragmatic experimenting and field testing and for a mutual tolerance of error. These needed qualities are inherent in science, so explaining them to nonscientists is a necessary first task in their joint venture. In this respect, the book reaches out to public policy-makers. But it serves another audience as well, an audience of scientists and engineers whose task it is to engage those nonscientists in asking the right questions. For this last group of readers, Branscomb offers an engaging model of how to go about the task. □

A. D. Van Nostrand is in the Library, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.



These alevins compete with fully grown specimens for photographic space in Adam Lewis's colourful and straightforward *Salmon of the Pacific* (Sasquatch Books, \$14.95). The author, a fisheries biologist, follows the famous Sockeye run on the Adams River in British Columbia to trace the fish's intriguing cycle of migration and reproduction.

The major evolutionary transitions

Eörs Szathmáry & John Maynard Smith

There is no theoretical reason to expect evolutionary lineages to increase in complexity with time, and no empirical evidence that they do so. Nevertheless, eukaryotic cells are more complex than prokaryotic ones, animals and plants are more complex than protists, and so on. This increase in complexity may have been achieved as a result of a series of major evolutionary transitions. These involved changes in the way information is stored and transmitted.

THE major evolutionary transitions¹ are listed in Table 1. There are common features that recur in many of the transitions: (1) Entities that were capable of independent replication before the transition can only replicate as parts of a larger unit after it. For example, free-living bacteria evolved into organelles². (2) The division of labour: as Smith³ pointed out, increased efficiency can result from task specialization (for a comprehensive review of this subject in the classical literature, see ref. 4). For example, in ribo-organisms nucleic acids played two roles, as genetic material and enzymes, whereas today most enzymes are proteins. (3) There have been changes in language, information storage and transmission. Examples include the origin of the genetic code, of sexual reproduction, of epigenetic inheritance and of human language.

Complexity

There is no generally accepted measure of biological complexity. Two possible candidates are the number of protein-coding genes, and the richness and variety of morphology and behaviour. Table 2 shows the sizes of the coding regions of various organisms⁵. The trend is fairly robust: eukaryotes have a larger coding genome than prokaryotes, higher plants and invertebrates have a larger genome than protists, and vertebrates a larger genome than invertebrates. The last observation is puzzling: perhaps the nervous system of vertebrates requires the extra genetic information. Unfortunately, the data do not tell us much about structural or functional complexity, because we do not know the mapping between genotype and phenotype.

Bonner⁶ measures complexity in terms of the variety of behaviour. For example, the emergence of humans depended on a greater behavioural variety. The point need not be confined to ethology: complexity increases with the diversity of actions an organism can carry out. For example, phagocytosis is a complex behaviour that depends on the eukaryotic cytoskeleton: prokaryotes cannot do it. The number of cell types in an organism can be taken as a measure of its complexity. Unfortunately, it is hard to quantify this aspect of complexity, or to get beyond the common-sense, but rather boring, conclusion that complexity has indeed increased in some lineages.

It is more interesting to list the mechanisms whereby the quantity of genetic information can increase. The three main possibilities—duplication and divergence, symbiosis and epigenesis—are shown in Fig. 1.

Transition from independent replicators

In many of the transitions listed in Table 1 we find the common phenomenon that entities capable of independent replication before the transition can only replicate as parts of a larger whole afterwards. Examples include the origin of chromosomes; the origin of eukaryotes with symbiotically derived organelles; the origin of sex; the origin of multicellular organisms (the cells of animals, plants and fungi are descended from unicellular protists, each of which could survive on its own: today, they exist only as parts of larger organisms); and the origin of social groups. Note that the last two examples differ from the previous

ones: the cells of multicellular animals did not form the organism through a symbiosis of independent entities, but they consist of entities (the cells), the analogues of which do exist as independent forms. Thus, units of evolution at the higher level may either be analogous (multicellular organisms) or homologous (eukaryotes) to an 'ecosystem' of lower-level units.

Given this common feature of the major transitions, there is a common question we can ask of them. Why did natural selection, acting on entities at the lower level (replicating molecules, free-living prokaryotes, asexual protists, single cells, individual organisms), not disrupt integration at the higher level (chromosomes, eukaryotic cells, sexual species, multicellular organisms, societies)? The problem is not an imaginary one: there is a real danger that selection at the lower level will disrupt integration at the higher. Some examples are¹: (1) If Mendel's laws are rigorously obeyed, a gene can only increase its representation in future generations by ensuring the success of the cell in which it finds itself, and of the other genes in the cell. Hence Mendel's laws ensure the evolution of cooperative, or 'coadapted', genes. But the laws are broken, in meiotic drive⁷, and by transposable elements⁸. These are examples of the more general phenomenon of intragenomic conflict⁹. (2) A sexual population has an advantage, in rate of evolution, and in the elimination of harmful mutations, over an asexual one. But a parthenogenetic female has, in the short run, a twofold advantage over a sexual one, and parthenogens are not uncommon¹⁰. (3) A gene in a somatic cell of a plant might best ensure the transmission of replicas of itself by giving rise to a flower bud, even if this reduced the success of the whole plant. (4) A bee colony produces more reproductives if the workers raise the queen's offspring. But workers do lay eggs (which are unfertilized, and hence male)¹¹.

We cannot explain these transitions in terms of the ultimate benefits they conferred. For example, it may be that, in the long run, the most important difference between prokaryotes and eukaryotes is that the latter evolved a mechanism for chromosome segregation at cell division that permits DNA replication to start simultaneously at many origins, whereas prokaryotes have only a single origin of replication¹². At the very least, this was a necessary precondition for the subsequent increase in DNA content, without which complexity could not increase. But this is not the reason why the change occurred in the first place: the new segregation mechanism was forced on the early eukaryotes by the loss of a rigid cell wall, which plays a crucial role in the segregation of eubacterial chromosomes. Or, to take a second example, meiotic sex was an important preadaptation for the subsequent evolutionary radiation of the eukaryotes, but it could not have originated for that reason.

The transitions must be explained in terms of immediate selective advantage to individual replicators. We are committed to the gene-centred approach outlined by Williams¹³ and made still more explicit by Dawkins¹⁴. There is, in fact, one feature of the transitions listed in Table 1 that leads to this conclusion. At some point in the life cycle, there is only one copy, or very few copies, of the genetic material: consequently, there is a high degree of genetic relatedness between the units that combine in

TABLE 1 The major transitions¹

Replicating molecules to populations of molecules in compartments
Unlinked replicators to chromosomes
RNA as gene and enzyme to DNA and protein (genetic code)
Prokaryotes to eukaryotes
Asexual clones to sexual populations
Protists to animals, plants and fungi (cell differentiation)
Solitary individuals to colonies (non-reproductive castes)
Primate societies to human societies (language)

TABLE 2 Genome size and DNA content⁵

	Genome size (base pairs $\times 10^9$)	Coding DNA (%)
Bacterium (<i>E. coli</i>)	0.004	100
Yeast (<i>Saccharomyces</i>)	0.009	70
Nematode (<i>Caenorhabditis</i>)	0.09	25
Fruitfly (<i>Drosophila</i>)	0.18	33
Newt (<i>Triturus</i>)	19.0	1.5–4.5
Human	3.5	9–27
Lungfish (<i>Protopterus</i>)	140.0	0.4–1.2
Flowering plant (<i>Arabidopsis</i>)	0.2	31
Flowering plant (<i>Fritillaria</i>)	130.0	0.02

the higher organism. The importance of this general principle was first emphasized by Hamilton¹⁵ in his explanation of the evolution of social behaviour, but we believe it to be quite general. To give two other examples: multicellular organisms develop from a single fertilized egg, so that their cells are genetically identical, except for somatic mutation; most eukaryotes inherit their organelles from one parent only, so that the organelles in a single individual are almost always genetically identical^{16,17}. We think that a similar principle operated in the origin of the earliest cells^{18–20}; this example is discussed further in Box 1.

In several of the listed transitions, one is effectively dealing with a group of replicators: when does such a group qualify as an organism, or—viewed from the level of the component replicators—a ‘superorganism’? Wilson and Sober²¹ define a superorganism as a “collection of single creatures that together possess the functional organization implicit in the formal definition of organism”. They suggest that groups are superorganisms if they satisfy the following criteria²¹: the population consists of several groups; there is a difference between the groups in their contribution of progeny to the next generation (differential group fitness); variation of group fitness is due to heritable genetic variation; individuals within the group have the same fitness.

Obviously, the last criterion cannot hold at the time of origin itself; it is precisely this absence of between-individual, within-group selection that has to be explained. The real question is whether a mechanism, suppressing internal competition, will invade when rare. The answer is known to be yes for certain cases (Box 1). For such a mechanism to spread, selection between groups must be effective. It is most efficient if the following criteria are met²²: the number of groups must be much larger than that of the units within each group; there is no migration (horizontal transfer) between groups; each group has no more than one parental group.

The effect of these conditions is that there will be genetic differences between groups, but individuals in a single group will be similar. If so, the groups will be units of evolution²³ and will evolve by natural selection.

The principle of the small number of founders is important at the time of the transition. Two other processes—contingent irreversibility and central control—help to explain the maintenance of higher-level entities once they have arisen, although they are less relevant to the origin of such entities¹.

■ **Contingent irreversibility.** If an entity has replicated as part of a larger whole for a long time, it may have lost the capacity for independent replication that it once had, for accidental reasons that have little to do with the selective forces that led to the evolution of the higher-level entity in the first place. For example, mitochondria cannot resume independent existence, if only because most of their genes have been transferred to the nucleus; cancer cells may escape growth control, but have no independent future as protists; worker bees may lay male eggs, but cannot establish a new colony on their own.

The contingent nature of irreversibility is perhaps best illustrated by the reversion from sex to parthenogenesis. Mammals are never parthenogens, probably because, at some loci in some tissues, only the allele inherited from the father is active: hence, in an embryo with no father, some essential gene activities are missing. Gymnosperms are also never parthenogens, perhaps for a different reason: chloroplasts are transmitted in the pollen. Anamniote vertebrates, although they may be parthenogens, always require sperm from the males of another species to initiate development, perhaps because the sperm provides a centriole. The relevance of these sexual hangups—and there are many others—is to show how various and accidental are the reasons why reversal is difficult or impossible.

■ **Central control.** If a ‘selfish’ mutation occurs in a chromosomal gene, a suppressor mutation at any other locus in the genome would be favoured by selection. Hence the rest of the genome may win the contest, not because of any analogue of majority voting, but because of the large number of loci, and hence of possible suppressor mutations, that are available for each selfish mutation. It may be relevant that attempts to use driving chromosomes in biological control have so far failed because of the rapid evolution of suppression. It is in this sense that Leigh’s idea²⁴ of a “parliament of genes” should be understood.

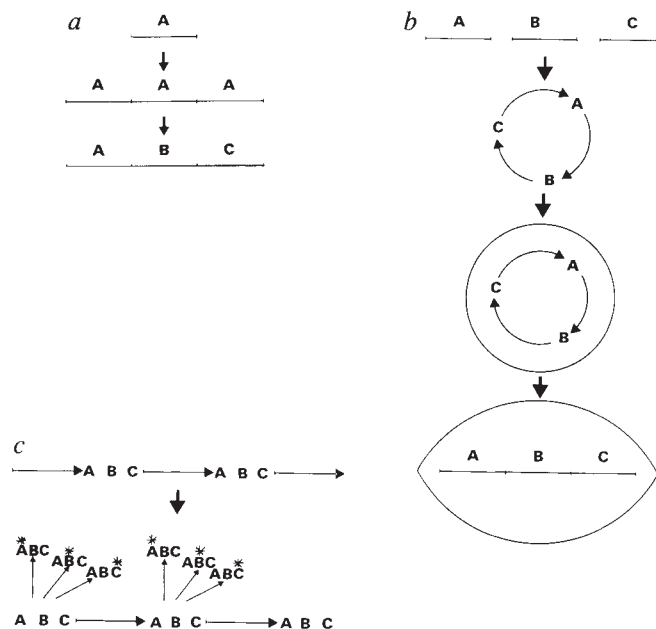
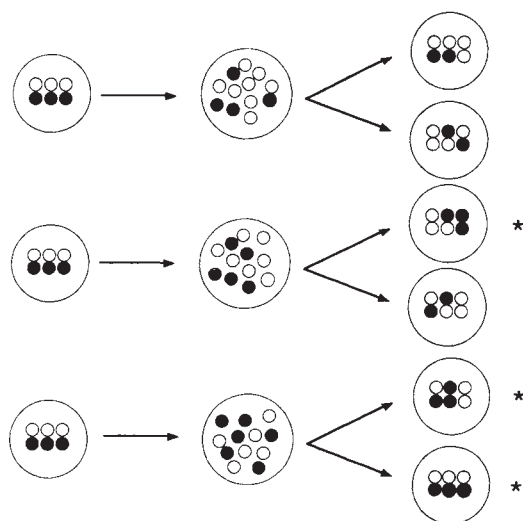


FIG. 1 Processes whereby the quantity of information can increase during evolution. a, Duplication followed by divergence: this is the main process whereby information increases between the major transitions. b, Symbiosis: the figure illustrates first a set of independent replicators, then a hypercycle⁶⁰, in which the replicators interact to form a stable ecological cycle, then the enclosure of the hypercycle in a compartment, and finally the physical linkage of replicators, so that when one replicates, all do. c, Epigenesis: A, B and C are different genes; asterisks indicate states of gene activity transmitted through cell division.

BOX 1 From naked genes to compartments to chromosomes

THE idea that the primordial genome must have consisted of unlinked genes comes from a paradox of Eigen. He demonstrated that the tolerable mutational load of a population sets an upper limit to the length of the genome: it is proportional to the reciprocal of the mutation rate per base per generation (the error threshold)⁶⁰. Early genomes could not have been much longer than a contemporary transfer RNA owing to the low fidelity of replication. A population in mutation–selection balance of such molecules (a so-called quasispecies⁶¹) cannot harbour a set of sufficiently dissimilar genes (to encode



The stochastic corrector model^{18–20}. Empty and filled circles represent two kinds of gene; the former have an average within-cell replicative advantage, but all genes replicate faster in cells with equal numbers of the two kinds. Stochastic processes (in replication and cell division) ensure the reappearance of cells, marked with asterisks, with the optimal gene composition, despite contrary within-cell selection.

different functions) in appreciable concentration. Therefore, a collection of unrelated and unlinked genes is needed, but they will compete with each other until only one gene (with the highest replication rate) survives, and in turn the genome is doomed to extinction; hence the paradox that long chromosomes are unstable because of excessive mutational load, and a set of small genes is unstable because of internal competition. As Eigen recognized, some functional coupling among the genes is necessary⁶⁰.

A possible resolution of Eigen's paradox is the 'stochastic corrector' model^{18–20} (figure). The assumptions are as follows: (1) Unlinked genes replicate in compartments. There are two types of gene, which replicate at different expected rates, so that there is between-gene selection within compartments. (2) Compartments reproduce when the number of genes they contain has doubled. The rate of compartment growth depends on the kinds of genes it contains, and is fastest when different kinds are present. Thus different genes contribute non-additively to compartment fitness, as emphasized in our discussion on the division of labour. (3) Replication is a stochastic process and assortment of genes into offspring compartments is random.

Given these conditions, there is efficient group selection at the compartment level. Despite internal competition, natural selection maintains a stable compartment distribution, and neither type of gene is lost. This is due to the stochastic processes generating variation between compartments, on which selection—between the proto-cells—can act.

The spread of chromosomes has also been analysed in the context of the stochastic corrector model. Chromosomes (linking complementing genes), when introduced in small numbers into some proto-cells of a simulated population, are established in the population, despite a twofold within-cell replicative disadvantage relative to individual genes⁶². The reasons are first, that linkage is a safeguard against internal competition of genes, as one cannot replicate without the other, and second, that it pays for a gene to sit on a chromosome, because it does not then run the risk of finding itself, after cell division, in a cell with low fitness due to the absence of its complementing partner. The chemical feasibility of this transition has been worked out⁶³, resting in part on the 'genomic tag' model⁶⁴, which argues for an ancestral role of tRNA-like structures as signals for RNA replication.

The division of labour

The most familiar examples of the advantages arising from a division of labour concern caste differentiation in the social insects. Bell²⁵ applied Adam Smith's ideas to a less familiar example, cell differentiation in the Volvocales²⁶. The specific name *Volvox weismannia* is a reminder that these algae are an excellent example of the segregation of germ line and soma. Most members of the order possess only a single cell type, which fulfils all vegetative and reproductive functions. In *Pleodorina*, there is partial division of labour: some cells start with vegetative functions, but later differentiate into gonidia (asexual propagules). The genus *Volvox* has a *bona fide* segregation of germ line and soma: germ cells are immotile in the centre of the spheroid, and somatic cells bear cilia but cannot divide. The benefit of differentiation has been demonstrated: colonies produce a larger bulk of smaller offspring than do single cells of a similar size.

One precondition for the division of labour in the Volvocales is that motile cells cannot divide, and mitosing cells cannot move, because the same organelles are used either as basal bodies or as centrioles²⁷. A similar argument was used by Buss²⁸ for the flagellated blastulae of the lower metazoa.

Some other cases in which a division of labour is evident are^{1,29}: (1) The evolution of many specific enzymes from a set of multifunctional low-efficiency enzymes. If the first protocells were equipped with a few multifunctional enzymes, more efficient enzymes could evolve only by duplication and divergence³⁰. (2) In the RNA world, RNA served both as genetic material and catalyst: today, DNA is the genetic material, and most enzymes are proteins³¹. (3) In prokaryotes there is a single cell

compartment, whereas in eukaryotes the genetic nucleus and metabolic cytoplasm are separated, and additional organelles have evolved, some recruited symbiotically². (4) In sexual populations, isogamy has repeatedly evolved to anisogamy, with differentiated sperm and ova¹⁰. (5) Hermaphrodites are replaced by separate sexes: the most convincing explanation for why some organisms are hermaphrodite and some dioecious is in terms of the advantages of a division of labour³².

If cooperation is to evolve, non-additive, or synergistic, fitness interactions are needed. If two or more cooperating individuals can achieve something that a similar number of isolated individuals cannot, the preconditions exist: the image to bear in mind is that two men, each with one oar, can propel a boat, but one man with one oar will row in circles¹. But the dangers of intragenomic conflict remain: both relatedness and synergistic fitness interactions are likely to be needed.

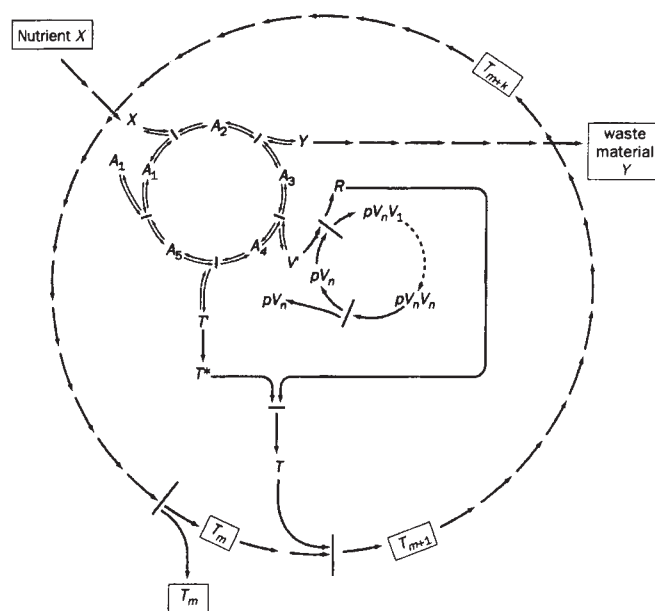
The evolution of heredity

Heredity means that like begets like: it requires some means whereby information can be transmitted. A crucial distinction is between systems of 'limited heredity', in which only a few distinct states can be transmitted, and systems of 'unlimited heredity', capable of transmitting an indefinitely large number of messages. We suggest the following stages¹.

(1) The origin of simple autocatalytic systems with limited heredity^{33–35}. Autocatalysis, whereby a single molecule gives rise to two molecules of the same kind, is essential for growth, but does not by itself imply heredity, which requires that, if the nature of the initial molecule is changed, two molecules of the new kind are produced. Several authors have suggested that

BOX 2 The chemoton as a sensible protocell model and its importance in explaining the first major transitions

THE basic model of the chemoton^{34,54,65} (figure) consists of three subsystems: the metabolic 'engine', which is an autocatalytic cycle; a self-replicating template macromolecule; and a bilayer membrane. The autocatalytic cycle produces the building blocks of the two other subsystems as well, at the expense of energy and material difference between X and Y . The condensation byproduct R serves as a stringent stoichiometric coupling between template polycondensation and membrane growth. It is easy to see that the whole system grows in synchrony. Division is a more tricky problem. There are calculations (which need to be verified experimentally) showing that a chemical system with the described couplings would indeed undergo spontaneous fission into two offspring compartments, owing to the interplay between growth, osmotic relations and surface tension of the membrane^{66,67}. The two questions we will discuss in turn concern



The chemoton⁵⁴. The metabolic subsystem, with intermediates A_i , is an autocatalytic chemical cycle, consuming X as nutrient and producing Y as waste material; pV_n is a polymer of n molecules of V , which undergoes template replication; R is a condensation byproduct of this replication, needed to turn T^* into T , the membranogenic molecule; the symbol T_m represents a bilayer membrane composed of m units made of T molecules. It can be shown that such a system can grow and divide spontaneously.

the realistic feasibility of the subsystems and how our central themes manifest themselves in the origin and evolution of such a system.

A remarkable example of an autocatalytic network is Butlerow's formose 'reaction', synthesizing sugars out of formaldehyde with the catalytic aid of pre-existing sugars (see for example, ref. 68). Other, still non-enzymatic, cycles were suggested as variations on the theme of the reductive citric acid cycle⁶⁹, as yet without experimental evidence. The origin of RNA-like self-replicating molecules is still a problem, as discussed in the main text (compare ref. 31). It is worth emphasizing the special kinetic effects within the chemoton, however: replication occurs only upon reaching a certain threshold concentration of V within the compartment, and this happens only once during the protocell cycle³⁴. The autocatalytic formation of membranes without enzymes is now proven⁷⁰. Experiments should now concentrate on the division mechanism.

As to our common themes, the following considerations are worth noting.

Complexity. Nobody thinks that a simple cycle in its drawn form could be realistic. Inevitably, one should have a network. The increase in complexity of such a network is an open problem. Chemical symbiosis⁷¹, or the grafting of novel extensions onto the pre-existing network³⁵ could have resulted in heritable, 'macroevolutionary' changes in the system. Apart from this, it is a family of templates, arising by mutation, duplication and divergence that can lead to a complex set of templates, which make use of digital information, first in the form of ribozymes catalysing steps of the metabolic network, and later as protein-coding genes.

Division of labour. It is common that symbiotic partners provide complementary metabolic 'toolkits' for the new unit². The chemoton's subsystems serve exactly such complementary roles, and their unit can be regarded as a very special case of chemical symbiosis. It is obvious that the membrane itself would be an inferior metabolic subsystem, and that templates could provide only leaky boundaries, so indeed there is an advantage in the union of specialized subsystems: molecular 'jacks-of-all-trades' are replaced by 'masters'.

Competition of replicators. The organization of the basic chemoton model is such that the complementary subsystems cannot get "out of phase"^{34,54}. This is not so within subsystems, most notably for the digital information carriers^{19,20}. Through microevolution, selfish mutants can arise. It is the stochastic corrector principle (Box 1) that can prevent the system as a whole from deteriorating.

Heredity. Two of the subsystems (the membrane and the metabolic cycle) carry only analog information⁷² and are at the most limited hereditary replicators. It is the realistic versions of the pV_n that can provide the system with unlimited heredity, because of their digital information⁷². First this is used for ribozymic activity, then in translation.

autocatalytic networks could have this property, but at best they could display limited heredity, with only a few molecular types able to reproduce themselves.

(2) The origin of polynucleotide-like molecules, providing unlimited heredity³¹. This transition has proven surprisingly difficult to explain. Even if one assumes the presence of all the necessary chemical constituents, there are severe obstacles to continued replication, such as enantiomeric cross-inhibition (mirror-image building blocks pair but do not form a covalent link with the growing chain³⁶) and non-separation of template and replica due to the many hydrogen bonds formed between them. Oligonucleotide replication is a possible intermediate stage leading to that of polynucleotides. Most such experiments use chemical analogues of oligonucleotides (such as the first successful attempt by von Kiedrowski³⁷). The short length allows for the spontaneous dissociation ('melting') of template and copy, so ongoing replication is possible, although an increased concentration is unfavourable for the latter because two complete strands find each other more readily. This results in a parabolic (subexponential) growth of such replicators³⁷, which in a competitive situation leads to a stable "survival of everybody"^{38,39} (see also ref. 40 for review). "Survival of the fittest" needs an exponential growth tendency³⁹ and therefore more efficient strand

separation. The latter could have been accomplished by RNAs with replicase function.

(3) The origin of the genetic code in the context of the RNA world, before translation. The essence of the code is that specific amino acids should be attached to specific oligonucleotides: today, it depends on the attachment of amino acids to transfer RNA molecules. Several workers in the field have realized that translation and coding are difficult to evolve simultaneously. One way to avoid such an evolutionary trap is preadaptation: rudiments of a complex adaptation may have evolved by selection for some other function. Concrete versions of this idea suggest that aminoacylation helped the replication of RNA, or that peptide-specific ribosomes with an internal message antedate general protein synthesis using external templates (messenger RNA). The idea that we favour is that amino acids were used as coenzymes of ribozymes, and were equipped with unambiguous trinucleotide handles, enabling them to bind to a ribozyme by base pairing⁴¹. Such handles would enable the same amino acid to be used by several ribozymes. Each new amino acid that acquired a specific handle would increase the enzymatic versatility of the organism, so that the difficulty of a complete adaptation being acquired in a single step largely disappears.

(4) The origin of translation and encoded protein synthesis. The details of this transition are discussed in ref. 1.

(5) The replacement of RNA by DNA as the genetic material could well have happened before the origin of genetic code, because it is chemically a much less complicated transition⁴². The primary selective force for this may have been the increased chemical stability of thymine (as opposed to uracil) and deoxyribose (as opposed to ribose)⁴³. The usual argument that there is no mismatch repair or repair of damage in RNA misses the point that, chemically, all these processes would be feasible in double-stranded RNA.

(6) The emergence of hereditary regulative states in prokaryotes and simple eukaryotes. Already in prokaryotes, patterns of methylation are transmitted through cell division and can be responsible for states of phenotypic differentiation. Thus there is a dual inheritance system, in which heredity depends either on differences in DNA sequence or on transmissible states of gene activation^{44,45}. Such a system is crucial for the development of animals and plants, but what selective forces were responsible for its evolution in single-celled organisms? Jablonka suggests that it was the need for protists to adapt to regular changes in the environment, the timescale of which was too large in comparison with generation times, and too small relative to the time required for typical evolutionary changes⁴⁴. In this case, a heritable mark M1 on some gene could have been beneficial in environment E1, and an alternative mark M2 (maybe simply the lack of M1) could have been beneficial in environment E2. An alternative suggestion is that morphological and physiological adaptations of sexual protists could have been preadaptations for simple forms of multicellularity, as alternative phenotypes, specific cell adhesion, cell-to-cell signalling and cell-division arrest play a crucial role in both⁴⁶.

(7) The evolution of epigenetic inheritance with unlimited heredity: the emergence of animals, plants and fungi. The transition to multicellular organisms with many kinds of differentiated cells occurred on three occasions, suggesting that it may not have been particularly difficult. This would be explained if the main cellular novelty required was an epigenetic inheritance system, as this existed already in protists. If so, the emergence and radiation of the metazoa had to wait only for suitable environmental conditions⁴⁷.

(8) The emergence of proto-language in *Homo erectus*—a cultural inheritance system with limited potential in which, because of the absence of grammar, only certain types of statement can be made⁴⁸.

(9) The emergence of human language with a universal grammar⁴⁹ and unlimited semantic representation⁵⁰. Grammar enables a speaker with a finite vocabulary to convey an indefinitely large number of meanings, just as the genetic code enables DNA to specify an indefinitely large number of proteins. We accept Chomsky's argument that grammatical competence is unique, both in the sense of being peculiar to humans, and of being special to language, and not merely an aspect of general learning ability. But we are puzzled by the reluctance of many linguists, including Chomsky himself, to think about the evolution of this competence. The objection takes the form of asserting, not only that human language is different in kind from animal communication, but that no intermediate is possible between the two.

It is argued that any rudimentary form of grammar would not allow one to generate some types of sentence. This is true but irrelevant: by analogy, it is better to have some light-sensitive cells than none at all; a perfect eye is not the only useful solution to the problem⁵¹.

It is in fact rather easy to think of intermediates between protolanguage and true language. There remains the question of the evolutionary origin of grammatical novelties. It is reasonable to assume that this happened by genetic assimilation⁵¹, new rules being made up by individuals as non-genetic innovations, then learnt by members of the community, then hard-wired into

the 'language organ' subsequently. It has been demonstrated that learning and selection can lead to such an assimilation in extreme cases when the latter alone could not get anywhere⁵²: learning can transform an initially flat fitness landscape with a needle-like peak into a well-behaved Fujiyama-like one.

Perhaps the most convincing evidence both for the belief that grammatical competence is to some degree independent of general learning ability, and for the possibility of functional intermediates between no grammar and perfect grammar, comes from studies of hereditary variation in linguistic competence. One remarkable case involves a family in which a so-called feature-blind dysphasia seems to be inherited in a mendelian fashion, a single dominant gene being responsible⁵³. Members cannot automatically generate plurals and past tense. Although they understand the meaning of plural and past perfectly well, they have to learn each new case anew: *paint* and *painted*, *book* and *books* must be learned separately (in the case of exceptions such as *go* and *went*, we must do the same). To be sure, this is not a genetical violation of one of Chomsky's rules, but it demonstrates that there can be something useful between perfect grammar and protolanguage: it also holds out the hope that we will in future be able to dissect language genetically, as we are today dissecting development.

Constructive evolution

Although some of the key intermediate stages of evolution seem to have vanished, their experimental recreation could teach us a lot. Examples include:

- The *de novo* synthesis of a living chemical system, such as the chemoton⁵⁴ (see Box 2 for a summary of this idea, and how several of our discussed points integrate into a unified picture at a certain level of organization).

- *In vitro* construction of a truly self-replicating RNA.

- *In vitro* generation of ribozymes, using amplification and selection by affinity chromatography^{19,55}. The first such example has been given⁵⁶. In a similar vein, the generation of RNA molecules of importance for primordial coding and translation, essentially by the same protocol, should provide us with useful information about feasible scenarios of the code's origin^{19,55}. Recently Famulok reported the *in vitro* selection of an RNA binding ornithine and citrulline⁵⁷. Similar tests using proteinogenic amino acids would be welcome.

- The establishment of artificial symbioses should help to clarify several aspects of some of the transitions. A first example is Jeon's bacteria, originally parasitizing an amoeba, which became obligatorily dependent on these bacteria later⁵⁸.

- Finally, recreation of extant species or forms may be in certain cases possible. The recreation of a fossil fern species from genomes of extant polyploids is a remarkable example⁵⁹.

Conclusions

A central idea in contemporary biology is that of information. Developmental biology can be seen as the study of how information in the genome is translated into adult structure, and evolutionary biology of how the information came to be there in the first place. Our excuse for writing an article concerning topics as diverse as the origins of genes, of cells and of language is that all are concerned with the storage and transmission of information. The article is more an agenda for future research than a summary of what is known. But there is sufficient formal similarity between the various transitions to hold out the hope that progress in understanding any one of them will help to illuminate others. □

Eörs Szathmáry is in the Collegium Budapest (Institute for Advanced Study), Szentháromság u. 2., H-1014 Budapest and Ecological Modelling Research Group, Department of Plant Taxonomy and Ecology, Eötvös University, Ludovika tér 2, H-1083 Budapest, Hungary; John Maynard Smith is at the School of Biological Sciences, The University of Sussex, Brighton BN1 9QG, UK.

1. Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman, Oxford, 1995).
2. Margulis, L. *Symbiosis in Cell Evolution* (Freeman, San Francisco, 1981).
3. Smith, A. *The Wealth of Nations* (1777).
4. Rensch, B. *Evolution above the Species Level* (Wiley, New York, 1966).
5. Cavalier-Smith, T. (ed.) *The Evolution of Genome Size* (Wiley, Chichester, 1985).
6. Bonner, J. T. *The Evolution of Complexity by Means of Natural Selection* (Princeton University Press, Princeton, NJ, 1988).
7. Crow, J. F. *BioEssays* **13**, 305–312 (1991).
8. Charlesworth, B., Sniegowski, P. & Stephan, W. *Nature* **371**, 215–220 (1994).
9. Hurst, L. *Proc. R. Soc. Lond. B* **248**, 135–140 (1992).
10. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, Cambridge, 1978).
11. Wilson, E. O. *Sociobiology: The New Synthesis* (Belknap, Harvard University Press, Cambridge, Massachusetts, 1975).
12. Cavalier-Smith, T. *Ann. N. Y. Acad. Sci.* **503**, 17–54 (1987).
13. Williams, G. C. *Adaptation and Natural Selection* (Princeton University Press, Princeton, NJ, 1966).
14. Dawkins, R. *The Selfish Gene* (Oxford University Press, Oxford, 1976).
15. Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
16. Hurst, L. D. & Hamilton, W. D. *Proc. R. Soc. Lond. B* **247**, 189–194 (1992).
17. Hutson, V. & Law, R. *Proc. R. Soc. Lond. B* **263**, 43–51 (1993).
18. Szathmáry, E. & Demeter, L. *J. theor. Biol.* **128**, 463–486 (1987).
19. Szathmáry, E. *Oxf. Surv. Evol. Biol.* **6**, 169–205 (1989).
20. Szathmáry, E. *Trends Ecol. Evol.* **4**, 200–204 (1989).
21. Wilson, D. S. & Sober, E. *J. theor. Biol.* **136**, 337–356 (1989).
22. Leigh, E. G. *Trends Ecol. Evol.* **6**, 257–262 (1991).
23. Maynard Smith, J. *Proc. R. Soc. Lond. B* **219**, 315–325 (1983).
24. Leigh, E. G. *Adaptation and Diversity* (Freeman, Cooper and Co., San Francisco, 1971).
25. Bell, G. in *The Origin and Early Evolution of Sex* (eds Halvorson, H. O. & Mornoy, A.) 221–256 (Liss, New York, 1985).
26. Kirk, D. L. *Trends Genet.* **4**, 32–36 (1988).
27. Koufopanou, V. *Am. Nat.* **143**, 907–931 (1994).
28. Buss, L. *The Evolution of Individuality* (Princeton University Press, Princeton, NJ, 1987).
29. Molnár, I. *Abstr. botanica (Budapest)* **17**, 207–224 (1993).
30. Kacser, H. & Beeby, R. *J. molec. Evol.* **20**, 38–51 (1984).
31. Joyce, G. *Nature* **338**, 217–224 (1989).
32. Charnov, E., Maynard Smith, J. & Bull, J. *J. Nature* **263**, 125–126 (1976).
33. Ycas, M. *Proc. natn. Acad. Sci. U.S.A.* **41**, 714–716 (1955).
34. Gánti, T. *A Theory of Biochemical Supersystems* (Akadémiai Kiadó, Budapest and University Park Press, Baltimore, 1979).
35. Wächtershäuser, G. *Microbiol. Rev.* **52**, 452–484 (1988).
36. Joyce, G. F., Schwartz, A. W., Orgel, L. E. & Miller, L. S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4398–4402 (1987).
37. von Kiedrowski, G. *Angew. Chem. Inter. Ed.* **25**, 923–935 (1986).
38. Szathmáry, E. & Gladkih, I. *J. theor. Biol.* **138**, 55–58 (1989).
39. Szathmáry, E. *Trends Ecol. Evol.* **6**, 366–370 (1991).
40. von Kiedrowski, G. *Bioorg. Chem. Frontiers* **3**, 113–146 (1993).
41. Szathmáry, E. *Proc. natn. Acad. Sci. U.S.A.* **90**, 9916–9920 (1993).
42. Benner, S. A. et al. *Cold Spring Harbor Symp. quant. Biol.* **52**, 56–63 (1987).
43. Lazcano, A., Guerrero, R., Margulis, L. & Oró, J. *J. molec. Evol.* **27**, 283–290 (1988).
44. Jablonka, E. *J. theor. Biol.* **170**, 301–309 (1994).
45. Jablonka, E. & Lamb, M. J. *Epigenetic Inheritance and Evolution: The Lamarckian Dimension* (Oxford University Press, Oxford, 1995).
46. Szathmáry, E. *J. theor. Biol.* **169**, 125–132 (1994).
47. Wolpert, L. *Biol. J. Linn. Soc.* **39**, 109–124 (1990).
48. Bickerton, D. *Language and Species* (University of Chicago Press, Chicago, 1990).
49. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA, 1965).
50. Pinker, S. *The Language Instinct—The New Science of Language and Mind* (Penguin, London, 1994).
51. Pinker, S. & Bloom, P. *Behav. Brain. Sci.* **13**, 707–784 (1990).
52. Hinton, G. E. & Nowlan, S. *J. Complex Syst.* **1**, 495–502 (1987).
53. Gopnik, M. *Nature* **344**, 715 (1990).
54. Gánti, T. *The Principle of Life* (OMIKK, Budapest, 1987).
55. Szathmáry, E. *Nature* **344**, 115 (1990).
56. Prudent, J. R., Uno, T. & Schultz, P. G. *Science* **264**, 1924–1927 (1994).
57. Famulok, M. *J. Am. chem. Soc.* (in the press).
58. Jeon, K. W. in *Symbiosis as a Source of Evolutionary Innovation* (eds Margulis, L. & Fester, R.) 118–131 (MIT Press, Cambridge, MA, 1991).
59. Rasbach, H., Rasbach, K., Reichstein, J. J. & Vida, G. *Ber. Bayer. Bot. Ges.* **50**, 23–27 (1979).
60. Eigen, M. *Naturwissenschaften* **58**, 465–523 (1971).
61. Eigen, M. & Schuster, P. *Naturwissenschaften* **64**, 541–565 (1977).
62. Maynard Smith, J. & Szathmáry, E. *J. theor. Biol.* **164**, 437–446 (1993).
63. Szathmáry, E. & Maynard Smith, J. *J. theor. Biol.* **164**, 447–454 (1993).
64. Weiner, M. & Maizels, N. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6729–6734 (1994).
65. Gánti, T. *BioSystems* **7**, 189–195 (1975).
66. Koch, A. L. *J. molec. Evol.* **21**, 270–277 (1985).
67. Tarumi, K. & Schwegler, H. *Bull. Math. Biol.* **47**, 307–320 (1987).
68. Cairns-Smith, A. G. & Walker, G. L. *BioSystems* **5**, 173–186 (1974).
69. Wächtershäuser, G. *Progr. Biophys. molec. Biol.* **58**, 85–201 (1992).
70. Bachmann, P. A., Luigi, P. L. & Lang, J. *Nature* **357**, 57–59 (1992).
71. King, G. A. M. *BioSystems* **13**, 23–45 (1980).
72. Wächtershäuser, G. *Proc. natn. Acad. Sci. U.S.A.* **91**, 4283–4287 (1994).

ACKNOWLEDGEMENTS. This work has partly been supported by the Hungarian National Scientific Research Fund (OTKA).

ARTICLES

A mechanism for decoupling within the oceanic lithosphere revealed in the Troodos ophiolite

Susan M. Agar^{*} & Kim D. Klitgord[†]

^{*} Northwestern University, Evanston, Illinois 60208, USA

[†] US Geological Survey, Menlo Park, California 94025, USA

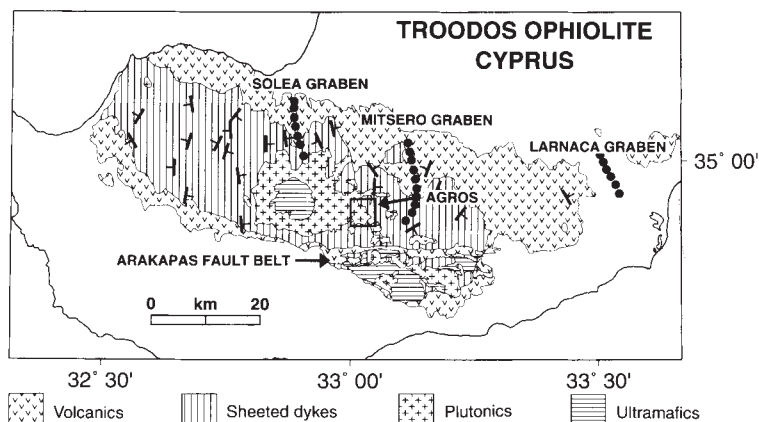
Contrasting kinematic histories recorded in the sheeted dykes and underlying plutonic rocks of the Troodos ophiolite provide a new perspective on the mechanical evolution of oceanic spreading centres. The kinematic framework of the decoupling zone that partitions deformation between the sheeted dykes and plutonics contrasts with low-angle detachment models for slow-spreading ridges based on continental-rift analogues. A model for the generation of multiple, horizontal decoupling horizons, linked by planar normal faults, demonstrates new possibilities for the kinematic and rheological significance of seismic reflectors in oceanic lithosphere.

IMPROVED acoustic images of the oceanic lithosphere have spawned new hypotheses on its architecture and evolution^{1–3}. Sub-horizontal reflectors have been interpreted as lithological transitions, metamorphic fronts, the roofs of active or fossil magma chambers and zones of magmatic underplating^{4–10}. Dipping reflectors have been equated with magma chamber margins¹¹ and faults^{1,5,10}. Some of these 'faults' appear to shallow at depth into sub-horizontal reflectors and have been interpreted as low-angle detachment surfaces either within seismic layer 3 (the lower crust) or along the Moho^{1,5,10}. Mechanical models of sea-floor spreading centres have attempted to integrate these subsurface features with fault geometries and kinematics interpreted from sea-floor morphology^{1,3,12}. Recognized limita-

tions to such an integration are the sparse subsurface features that can be traced to the sea floor, the ambiguous origins of reflectors and the lack of constraints on fault displacement vectors and timing. Consequently, spreading-centre models have incorporated geometries and kinematics established for continental extensional fault systems^{1,13–20}. Low-angle detachments, which arise at brittle–ductile transitions, have been proposed as a mechanism for the exhumation of deep crust and mantle sections on rift flanks during amagmatic cycles^{15–20}. The test for such models depends on critical evaluations of the distribution, timing and magnitude of deformation in the lithosphere.

Ophiolites provide the three-dimensional exposures and age relations needed to complement seismic images and drill holes

FIG. 1. Geological setting of the Agros window, Troodos ophiolite, Cyprus⁴⁰. The Troodos ophiolite is an obducted fragment of Cretaceous "oceanic crust" that is exceptional in its lack of structural and metamorphic overprinting associated with emplacement^{21,23,25,40}. The complex includes an outer rim of extrusives, a sheeted-dyke complex and a central zone of plutonics (mainly gabbros) with an inner core of ultramafics. This inner core is exposed by a Late Tertiary serpentine diapir on Mt Olympus which has imposed only minor tilting on the Troodos complex^{27,30,32}. The general north-south trend of dykes within the sheeted-dyke complex indicates that Troodos may represent an east-west cross-axis section of a spreading centre^{21,27}. The east-west-trending Arakapas fault belt has been interpreted as the northern edge of a transform fault zone⁴³⁻⁴⁵. Three north-trending grabens have been identified within the sheeted-dyke complex (the Solea, Mitsero and Larnaca grabens) and have formed the basis for several models of rift geometries, ridge jumps and propagating spreading centres^{27,30,40}. Box indicates the location of the Agros region where resistant sheeted dykes form high-standing ridges surrounding erosional valleys in the plutonics.



at present-day spreading centres²⁰⁻²⁵. Although geochemical signatures indicate that ophiolites formed in a supra-subduction zone environment²⁶, the lithospheric architecture closely resembles that of oceanic lithosphere. For this reason, ophiolites provide a valuable alternative to continental-rift analogues to investigate sea-floor spreading mechanisms.

In an erosional window in the Troodos ophiolite, Cyprus (Fig. 1), we have discovered evidence for deformation partitioning across a major sub-horizontal decoupling surface in the transition zone between the sheeted dykes and underlying plutonic rocks. We observe contrasting kinematic histories of the sheeted dykes and plutonics, which have not been documented in previous studies of low-angle faults in the Troodos ophiolite²⁷⁻³³, in other ophiolites or at active spreading centres. Previous ophiolite studies have identified hanging-wall deformation associated with low-angle faults but the kinematics of the footwall have not been addressed, in many cases owing to lack of exposure or overprinting by obduction-related deformation^{22,24,34}. Where adequate exposures exist the footwall has been interpreted as a static feature (except for isostatic adjustments), with no rotation occurring below low-angle faults localized at the dyke-pluton boundary³³. Some models have delineated low-angle detachment faults either within the lower plutonics or along the lower-crust/mantle boundary, proposing that the hanging-wall lithologies rotate as one coherent block¹⁸⁻²⁰. Our results provide new insight into the kinematic and rheological significance of low-angle faulting in the ocean crust, and the role of mid-crustal brittle failure in channelling hydrothermal fluid flow. We present a kinematic model for the generation of linked, rift-bounding faults and sub-horizontal decoupling surfaces during sea-floor spreading. Low-strain, but kinematically significant, intracrustal decoupling in the brittle regime at the dyke-pluton transition is inherent to this model. Our model contrasts with extension models involving a high-strain, low-angle shear zone that detaches along the brittle-ductile transition¹⁵⁻²⁰ and exhumes lower-crust or mantle sections in the footwall. The similar fault geometries that characterize these alternative models may not be readily distinguished in seismic records but they represent distinctly different kinematic histories.

The Agros window

The Agros window exposes a vertical section (up to 650 m high) through the sheeted-dyke complex into the underlying plutonic section. The igneous relations in this lithologically controlled depression support a multiple magma-chamber model for the Troodos ophiolite, in which individual intrusions, rising to different crustal levels, generated clusters of sheeted dykes^{21,35-38}. A consequence of this polyphase magmatism is the generation

of a gradational dyke-pluton boundary, where the sheeted-dyke complex is intruded by plutonic rocks and (less commonly) dykes intrude the plutonic rocks^{37,38}. In the Agros window, the proportion of dykes increases towards the top of the plutonic section (from <10% to ≤70% by area over a 250-m vertical section). The transition from the uppermost plutonic section (mainly isotropic gabbros) to the 100% sheeted-dyke complex is sharp, however, varying by less than 50 m in relief across the north wall of the Agros window. There is a resistant, northeast-trending ridge (dyke keel, Fig. 2) of diabase dykes (100 m thick) that intrudes the plutonics but does not penetrate the sheeted-dyke complex. This diabase ridge is similar to other diabase keels sourced by deeper-level plutonics³⁷.

The Agros window is located 5 km west and 8 km southeast, respectively, from the inferred axes of the Mitsero and Solea grabens (Fig. 1). Conflicting tectonic models^{21,27-32,39,40} identify three main grabens (Mitsero, Solea and Larnaca grabens) as possible spreading centres in the Troodos ophiolite. The Agros window crust could therefore have been generated on either the eastern or western flank of a spreading centre. Inferred complexities of the boundary between the Mitsero and Solea grabens have been exacerbated by the selection of different representative dyke orientations from the same original 1960s datasets^{38,41} in different publications, and contradictory structural data^{27,32,39,42}. Our detailed mapping, however, shows the remarkable consistency of the sheeted-dyke strikes and plutonic history in the Agros window with adjacent regions, several kilometres to the north, east and west^{25,30,37,38,41}. From this we infer that the Agros window is not an anomalous section (as might be assumed from speculative models for overlapping spreading centres³⁹) and that its structures are most consistent with a spreading centre located to the east. The Arakapas fault belt, interpreted as the northern margin of a transform fault zone, lies 8 km to the south where the prevalent northerly strikes of the sheeted-dyke complex rotate clockwise to nearly east-west orientations^{41,43-45}.

Structural framework

We identify three major structural domains within the Agros window by their location relative to a north-trending, east-dipping fault on the western side of the Agros window (Fig. 2, "western" fault). The sheeted dykes have been down-faulted on the western fault and juxtaposed with the plutonic section in the footwall. In the footwall of the western fault, a thin (30-m) section of mainly steeply east- and west-dipping, sheeted dykes and a transitional zone of dykes with gabbro screens (about 50 m thick) overlies the plutonics. Small dyke clusters and isolated dykes in the plutonics dip predominantly to the east, strik-

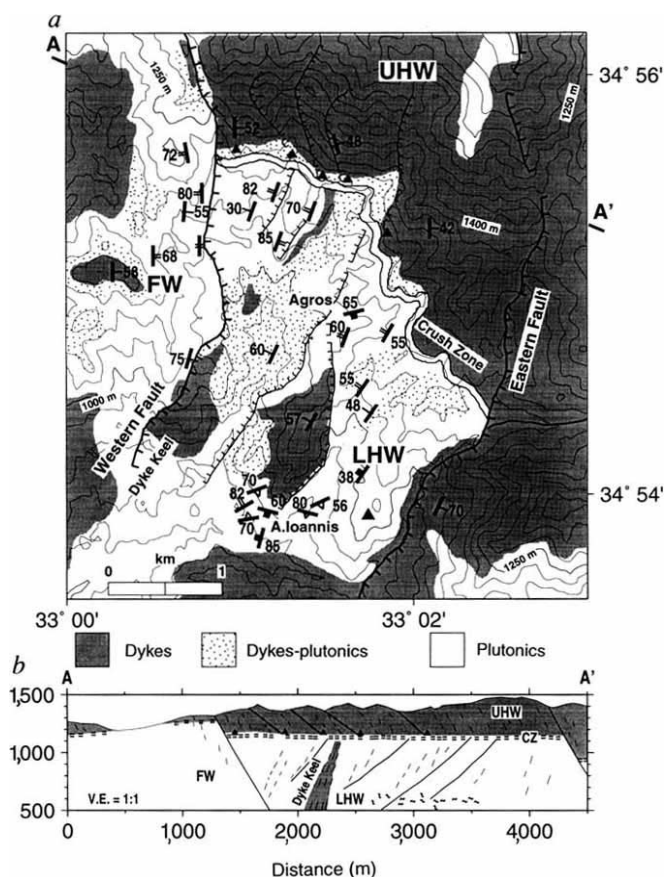


FIG. 2 a, Geological map of the Agros window. The lithological units are indicated by the shading: diabase dykes, dark shading; mixed dykes (variably altered basalt, diabase, plagiogranite, microgabbro) and gabbro (mainly gabbro-norite) (>30% dykes), stipple; plutonic section (gabbro-norite with olivine gabbro, minor wehrlite and plagiogranite), unshaded. Four distinct generations of dykes intrude the high-level gabbros (diabase, plagiogranite, basalt and pervasively altered, chloritized dykes) but the sheeted dykes comprise predominantly diabase. The uppermost plutonic section is dominated by varitextured, isotropic gabbro-norites but sparse magmatic foliations are present. This foliation may represent dyke root zones similar to those reported from the Oman ophiolite⁴³ or differential compaction of partially crystallized material caused by subsequent intrusions. Lower in the section there is an increased proportion of olivine gabbros with small regions of weakly defined layering close to Ayios Ioannis (too small to show at map scale). The two major east-dipping faults crossing the area are labelled Western fault and Eastern fault. Strike and dip marks (lines with single tick) show individual measurements of dykes in each structural domain, identified by their position relative to the western fault (FW, footwall domain; UHW, upper hanging-wall domain; LHW, lower hanging-wall domain). Fragmented mylonite zone orientations, lines with triangles; magmatic foliations, lines with double ticks; magmatic layering, lines with filled squares. The western fault downthrows the sheeted dykes by ~150 m. This estimate assumes that the thin carapace of mixed dykes and gabbros in the footwall of the western fault is the equivalent of the base of the dyke-pluton transition in the hanging wall. The northeast-trending dyke keel within the gabbroic section is indicated. The large triangle near Ayios Ioannis represents a major zone of hydrothermal alteration associated with an underlying pluton⁵¹. Smaller triangles just above the crush zone are smaller pods of hydrothermal alteration formed at the intersection of the domino faults in the upper hanging wall with the crush zone. b, Cross-section along A-A' in a, showing opposing dips of faults and dykes in the upper and lower hanging-wall domains of the Agros window. The transitional zone between the sheeted-dyke complex and gabbros is too thin to show. Schematic dyke margin orientations are shown by fine dashed lines. Squiggles, dismembered mylonite zones; CZ, crush zone; V.E., vertical exaggeration.

ing due north (Fig. 3a). The minimum vertical displacement on the western fault is ~150 m.

The upper hanging-wall domain comprises a 300-m-high section of the sheeted-dyke complex (Fig. 2). The faulted junction with the footwall is represented by a 50-m-wide zone of dense zeolite-veining with minor quartz and epidote. The dykes dip predominantly to the east (55° E) and strike NNW (Figs 2 and 3b). The majority of dykes are oriented subparallel to each other with only rare instances of cross-cutting dykes. At least four north-striking faults dissect the sheeted-dyke complex, each extending from the top of the ridge north of Agros down to the dyke-pluton transition (Fig. 2). These faults, spaced at 300–500 m intervals, are planar and dip 55° E. Kinematic indicators suggest oblique-normal displacements. Numerous minor faults with normal displacement components of less than 1 m have localized along dyke margins.

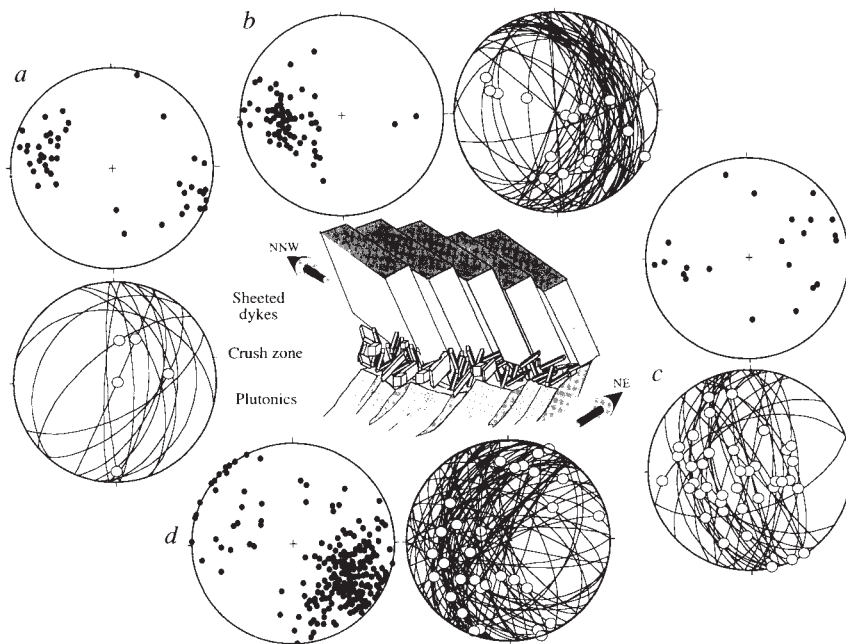
The lower hanging-wall domain contains polyphase gabbroic intrusions, intruded by numerous isolated dykes and the dyke keel. The dykes intruding the plutonics are predominantly west-dipping (about 60°) and strike NNE (Fig. 3d). Three, continuous (up to 2 km along strike) west-dipping fault zones dissect the plutonic sequence with numerous, minor (2–3-m offsets) west-dipping normal faults interspersed between them. Dense, distributed fracturing accommodates the transition from west dip of dykes and faults in the lower hanging wall to the predominantly east dip of structures in the footwall over a 500-m-wide zone. Most of the deformation in the plutonics is brittle. A sparse, steeply dipping syn-magmatic foliation (which shows no microstructural evidence for crystal-plastic deformation) occurs in some olivine gabbros but rarely develops in zones wider than 3–5 m. The foliations have variable dips but are generally moderately-to-steeply west-dipping in the lower hanging wall. Weakly defined layering is present in small areas in olivine gabbros near Ayios Ioannis and to the east of Agros (Fig. 2) but is also too discontinuous and disrupted by faulting to trace possible

magma-chamber boundaries. The only strong penetrative crystal-plastic deformation is represented by mylonitic shear zones (with widths of a few centimetres to more than 1 m) which deform gabbros in the Ayios Ioannis region⁴⁶. These variably oriented mylonite zones (Fig. 2) have been fragmented by subsequent faulting and cannot be restored to their original orientations.

The junction between the upper and lower hanging-wall domains must accommodate the abrupt flip in the dips of dykes and faults and the relative rotation of their strikes. This junction lies within the dyke-pluton transition, represented by a “crush zone” which has a thickness of 50–100 m (Figs 2 and 3c). The crush zone is a macroscopic breccia where distributed fracturing and locally penetrative cataclasis isolate and deform decimetre-scale blocks of plutonics and dykes. No penetrative crystal-plastic deformation is evident. The zone is sub-horizontal and can be traced for over 5 km across the northern wall of Agros valley. The planar faults in the sheeted dykes merge into the crush zone with no decrease in dip except for a limited area (of about 30 m lateral extent) immediately above the crush zone. The deformation in the crush zone has prevented the identification of individual intrusive bodies either side of the zone, but the close similarities in diabase and gabbro lithologies and their metamorphic grade above and below the crush suggest it accommodates a relatively small displacement. The crush zone is cut by the western fault. We have not identified the crush zone in the footwall but the erosion level there is close to the projected level of the crush zone within the dyke-pluton transition. The crush zone may be truncated at the eastern margin of the window by the continuation of a second north-striking, east-dipping fault that down faults the sheeted-dyke complex against the plutonic section to the north (see “eastern fault”, Fig. 2).

Timing and conditions of deformation. We have found no dykes crossing faults either in the upper or the lower hanging-wall domains or the crush zone. The rotation of fault-bound

FIG. 3 Equal-area stereonet showing poles to dyke margins and great circles for fault planes in the structural domains of the Agros window. The diagram at the centre of the figure shows the relationships between the east-dipping sheeted dykes, the crush zone and west-dipping dykes in plutonics in the hanging wall of an east-dipping, rift-bounding fault. Points on great circles represent the slicken-side pitch on fault planes. Oblique normal displacements were evident where the shear sense could be determined. *a*, Footwall domain (average poles: dykes 282° , 63° ; faults 277° , 55°). Note that both steeply east- and west-dipping dykes are present. *b*, Upper hanging-wall domain (average poles: dykes 265° , 44° ; faults 270° , 49.9°); *c*, crush zone; *d*, Lower hanging-wall domain (average poles: dykes 118° , 37° ; faults 110.4° , 49.9°). The average strike of the dykes in sheeted dykes and plutonics are separated by 33° . The average strike of the faults in the sheeted dykes and the plutonics are separated by 20.9° . The average dips of faults in the upper and lower hanging-wall domains are the same (40.1°), whereas the average dip of dykes in the sheeted dykes is 46° but in the plutonics the dyke dips average 53° . All generations of dykes are combined on these plots and are not separated by different symbols. The separate data for each generation show no evidence for a sequential decrease in dips with each successive generation of dykes, supporting our interpretation that most fault block rotations are post-magmatic.



“domino” blocks in the sheeted dykes and faulting in the plutonic section were therefore mainly post-magmatic, although a few rare cross-cutting dykes could indicate the late stages of magmatism during block rotation. The prevalent west dip of all four generations of dykes in the lower hanging-wall domain (Fig. 3d) also suggests that any tectonic rotation of this lower block was post-magmatic.

Fault-rocks in the upper hanging-wall domain indicate syn- and post-kinematic fluid flow associated with localized high-temperature alteration (350°C)⁶. Irregular pods (5–15 m wide) and veins of quartz, epidote and sulphides are found at the base of the sheeted-dyke complex (triangles in Fig. 3), directly above the crush zone in the hanging walls and fault zones of the four domino faults. Vertical offsets of the crush zone are not apparent and the domino faults do not penetrate below it. The crush zone therefore represents a likely accommodation zone that formed synchronously with the domino faulting (Fig. 3, centre). Fractures in the crush zone are sealed by syn-kinematic hydrothermal assemblages which closely resemble those in the sheeted dykes and are similar to sulphide deposits along oceanic spreading centres^{47–50}. The syn-kinematic, high-temperature alteration indicates that at least some of the faulting in the sheeted dykes and the crush zone occurred either on-axis or close to the spreading centre.

The crush zone does not cut the western fault and was therefore generated before or during the generation of the western fault. The west-dipping faults in the gabbros do not cut up through the crush zone and no west-dipping faults have been found in the overlying sheeted-dyke complex. This suggests that faulting in the lower hanging wall was also accommodated by the overlying crush zone. Hydrothermal alteration in the lower hanging-wall domain is characterized primarily by low-temperature phases (zeolites) with minor quartz and epidote; however, there are several discrete (1–10 m thick) west-dipping fault zones, containing quartz, epidote, sulphides and chlorite, which trend NNE from Ayios Ioannis (Fig. 2). The fault seal assemblages closely resemble those of fault breccias in the crush zone and the sheeted dykes, and are closely associated with the last generation of strongly altered dykes that cut through the plutonics. The Ayios Ioannis region is extensively hydrothermally

altered and has been interpreted as the roof zone of a late pluton⁵¹. At least three fault zones can be traced from Ayios Ioannis up to the crush zone, apparently having channelled hot fluids through (upwards and laterally) the plutonic section, away from a late-stage pluton. The Ayios Ioannis pluton appears to be the most likely source for high-temperature fluids that permeated the crush zone and sheeted dykes, suggesting that at least some faulting in the plutonic section was synchronous with the sheeted-dyke and crush-zone deformation.

The western fault zone adjacent to the lower hanging wall is pervasively altered to clays and zeolites and contains only minor, high-temperature hydrothermal phases. Syn-kinematic zeolites also seal several lower hanging-wall fault zones and some of the crush-zone breccias. Cross-cutting relations show that low-temperature phases are both cut by and cut high-temperature fracture-fill assemblages in the plutonics. It is therefore possible that some low-temperature alteration occurred at the same time as the high-temperature alteration, with high-temperature fluids restricted to major fault-controlled flow paths and low-temperature phases distributed more pervasively away from these channels.

A kinematic model for intracrustal decoupling

These observations suggest a kinematic model for the observed deformation partitioning across the dyke-pluton transition at a spreading centre. This model (Fig. 4a) incorporates a down-temperature path that could represent motion of the Agros crust away from the spreading centre or the waning of a magma chamber. The successive stages of the model are represented at increasing distances from the spreading axis.

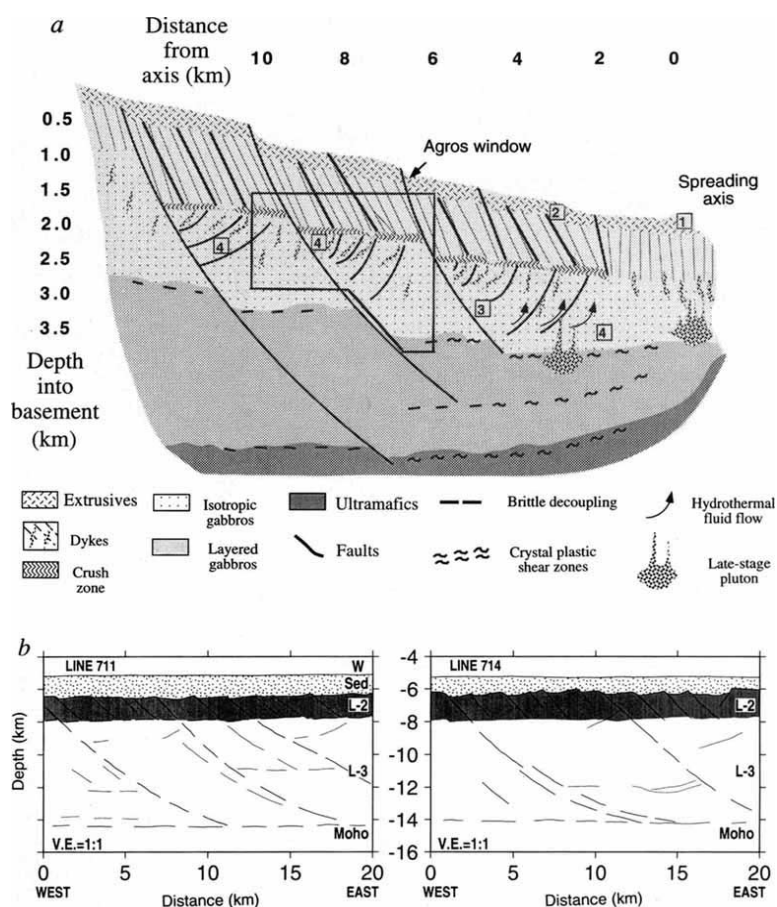
Stage 1. Within the axial zone, sea-floor spreading generated a coherent lid of steeply dipping dykes maintained by continuous dyke injection. This assumption is compatible with field observations of extensive sections of coherent packets of sheeted dykes^{34,52} and geophysical models^{6,7,12}. We recognize that some sheeted-dyke sections are highly incoherent both on the sea floor⁵³ and in the Troodos ophiolite, and may represent variations in magmatic activity or tectonic setting (see, for example, refs 45 and 54). Matching extension in the underlying plutonics at this stage was accommodated by a combination of near-axis

FIG. 4 a, Model for the structural evolution of a spreading centre inferred from the Agros window. Numbers in small boxes represent the different stages of the model discussed in the text. Although each stage is represented in a sequence during motion off-axis, the same sequence could occur on-axis during an episode of waning magmatism. Polygon indicates the approximate crustal location of exposures in the Agros window. The sea-floor relief in this model is taken from the AMAR region¹³ but a slow spreading regime is not necessarily assumed. The relatively small fault displacements on the western fault and in its hanging wall are consistent with an intermediate spreading regime. Our model assumes a moderate-to-slow spreading rate environment where the development of the crush zone is related to an episodic magmatic supply derived from small, transient magma chambers. The recent identification of small magma chambers and fluctuations in magma supply at fast-spreading ridges^{9,10} means that our model may be appropriate for these regimes as well. The spreading rate for the Troodos ophiolite has not been determined, but slow^{3,20,27,29} and intermediate-to-fast rates^{28,30} have been proposed. The assumption that dykes are generally intruded vertically is supported by observations from mid-ocean ridges⁵⁵. The different dips and strikes of dykes and interspersed faults in the upper and lower hanging walls indicate that some differential rotation must have occurred between these two distinct domains. b, Line drawings of migrated seismic-reflection lines 711 and 714 in the western Atlantic (based on Figs. 10a and 11b of ref. 11). These profiles are approximately normal to the spreading axis and image 155 Myr and 140 Myr oceanic crust, respectively. The faults penetrating to Moho appear to be rift-parallel faults dipping east, but the track spacing is too coarse to clearly define the fault geometries¹¹. The more closely spaced faults restricted to the upper-crustal unit also appear to dip eastwards towards the spreading axis, and may be analogous to the domino faults found at Agros. Note that the resolution of the seismic image in the upper crust is limited, and that these closely spaced faults are based on our interpretation of short sections of west-dipping basement and sporadic east-dipping reflectors penetrating the upper crust. This interpretation is only meant to illustrate the potential structural content available in recent seismic-reflection data sets. On both profiles the sedimentary layer (sed) is indicated by stipple pattern, seismic layer 2/3 boundary is at ~8 km depth and Moho is at ~14 km depth. The seismic layer 2/3 boundary is well defined on line 711 and would

or on-axis plutonism, off-axis dyke injections and crystal-plastic deformation^{46,55}.

Stage 2. Eastward tilting of the sheeted-dyke complex, accommodated by domino faults, fragmented 500-m-wide sections of the sheeted dykes from the initially coherent axial 'lid', generating the crush zone at the dyke-pluton transition. The brittle extension in the dykes may have accommodated a diminishing magmatic supply at the spreading axis. This zone of domino fault blocks was bound by east-dipping faults on the east and west margins. In the cooling thermal regime of waning axial magmatism, late, deep stocks (for example, Ayios Ioannis pluton) generated dyke keels which penetrated the overlying plutonics. Some of these dykes also penetrated the sheeted-dyke complex, producing the few cross-cutting dykes in the upper hanging-wall domain. Within a short (about 1 km) distance off-axis, extension continues to be accommodated at least in part by magmatism in the plutonic section while brittle failure was occurring in the sheeted dykes.

Stage 3. Some faults propagated down into the cooling plutonic complex. The western fault propagated through the crush zone, down-faulting the hanging-wall domains. In the Agros window we only know that the footwall moved up relative to the hanging wall but, by analogy with mid-ocean-ridge observations¹², this fault is likely to have been one of a series of normal faults accommodating uplift of the rift flank mountains (Fig. 4). In the absence of a continuing magma supply, extensional faults,



correspond to our crush zone. It is offset by small amounts (<100 m) on moderately-to-steeply dipping faults that penetrate into the lower crust. The seismic layer 2/3 boundary is poorly defined on line 714, but several of the upper-crustal faults seem to terminate within the upper 2 km below basement. The lower-crustal reflector at ~12 km depth terminates against the interpreted rift-parallel faults and could be one of the deeper decoupling surfaces.

antithetic to the western fault were generated below the crush zone, reactivating the crush zone and rotating the plutonic section to the west. The faulting and fracturing of the plutonic section provided pathways for high-temperature hydrothermal fluids from the late Ayios Ioannis stock, and may have assisted the intrusion of late-stage, strongly altered dykes in the plutonics. The crush zone provided a permeable conduit across the dyke-pluton transition, dispersing fluids up into existing fault zones in the overlying sheeted dykes. The low-temperature alteration phases around the western fault may represent a downflow zone for cold sea water⁵⁶ but these phases are also more pervasively distributed through the plutonic section.

Stage 4. Faults propagate down to Moho level as the lower crust cools into the brittle regime¹⁸. Cold sea-water influx and the waning of the Ayios Ioannis convection resulted in the continued precipitation of lower-temperature alteration phases during continued minor extensional faulting in the plutonic section, accommodated by the crush zone.

Apparent rotations about a vertical axis. This two-dimensional model does not account for the relative rotation of the strikes of the dykes between the upper and lower hanging-wall domains. The northerly strike of dykes is consistent with those surrounding the Agros window, suggesting that the gabbros have rotated beneath the strong 'lid' of sheeted dykes. The abrupt change in dyke orientations across the crush zone argues against their origin as primary, intrusive relations. The strike difference does

not necessarily require a rotation about a vertical axis. The dyke strike would change with rotation of the plutonics about a horizontal axis having a strike different from that of the sheeted dykes²⁹. The dyke orientations in both the footwall and the lower hanging wall of the Agros window maintain generally consistent strikes (north–WNW and NNE–northeast respectively) in the south of the area. Dyke orientations to the west of the area mainly indicate northerly strikes and east dips⁴¹. In a half-kilometre-wide zone adjacent to an east–west trending topographic lineament, south of Ayios Ioannis, several strike readings have northeast and ENE orientations. This narrow zone delineates the boundary of a new structural domain to the south that may be related to the southern Troodos transform zone. It is possible that the apparent clockwise rotation of the lower hanging wall may have been caused by shear displacements on the Arakapas fault to the south, synchronously with the extensional faulting outlined in our model. A similar mechanism has been proposed to explain the apparent clockwise rotation of the sheeted dykes southeast of Agros³⁰. Local rotations of the plutonic section associated with underlying mantle flow patterns^{23,24} or transfer zones linking rift-bounding normal faults¹² are possible, but untested, alternative explanations.

Broader implications of the decoupling model

Ocean-crust deformation models have been based primarily on inferences from sea-floor relief and rock samples^{3,12,15–21}, interpretations of seismic reflectors based on continental analogues¹ and mantle flow patterns²⁴. Tectonism along the dyke–pluton boundary has been reported from Troodos²⁷, other ophiolites³⁴ and sea-floor exposures^{53,57}, but until now comparisons of the kinematic frameworks of adjacent sheeted-dyke and plutonic sections have not been examined. The few structural and palaeomagnetic studies in the Troodos ophiolite have focused on the sheeted-dyke section and overlying extrusives^{27–32}.

The Agros structural geometries reveal vertical partitioning of deformation associated with rift-parallel faults during spreading-centre extension. The western and eastern faults at Agros (Fig. 2) are characteristic of the rift-parallel faults at about 2–5 km spacing found on mid-ocean ridges^{3,12}. We suggest that the 500-m spacing of planar faults within the sheeted dykes represents the down-dip continuation of more closely spaced faults reported from sea-floor spreading centres¹². The Agros crush zone isolates the sheeted-dyke rotations from the rotations of the underlying plutonics on west-dipping faults that are antithetic to the western fault. Discontinuous mylonite zones identified in the structurally lower parts of the Agros plutonic section may represent relicts of a high-temperature accommodation zone (decoupling surface) for these antithetic faults. We suggest that the mylonites have since been dismembered by moderate- to low-temperature faulting in the rift-flank mountains.

Previously described structures in the Troodos ophiolite bear some similarities to the crush zone and may have a similar origin to that proposed by our model. There is a brecciated zone within the dyke–pluton transition in the CY-4 borehole (7 km east of the Agros window) and several sheared contacts between the sheeted dykes and plutonics in the surrounding area^{37,39}. In the Solea graben (Fig. 1), the shallow-dipping Kakopetria detachment forms a sheared contact between the sheeted dykes and plutonics^{27,31,32}. No detailed accounts of the conditions of deformation indicated by fault rocks in the Kakopetria detachment have been published, but all of the reported structures associated with this shear zone are brittle and no high-temperature crystal-plasticity has been documented. Yet the Kakopetria detachment has been compared to continental and inferred oceanic detachment-fault zones^{27,30,32,39} where a sub-horizontal surface decouples at the brittle–ductile transition and high-temperature metamorphic rocks in the footwall are exhumed^{14,19}. The kilometre-scale displacements inherent to this type of detachment model and the accompanying down-temperature deformation paths during exhumation have not been found in the Kakopetria

detachment^{27,31,32}. The Akapnou Forest décollement in the Limassol Forest complex might also represent a deeper-level equivalent to the Agros crush zone, but its low-temperature, post-volcanic deformation has been used to infer an origin unrelated to sea-floor spreading processes³³. It may be possible to link the crush zone with other sub-horizontal faults localized along the dyke–pluton boundary. We think that the crush zone may have developed during periods of low magmatic supply. A discontinuous trace of the crush zone in Troodos may also be due to its partial assimilation by late-stage, shallow-level magmatism²⁵.

Evidence from seismic images of oceanic lithosphere^{5,7,10,11}, ODP boreholes^{46,58,59} and other ophiolites^{16,18,19,60} indicates that decoupling surfaces may exist at various depths in the oceanic lithosphere. A close examination of the reflector patterns on seismic images for seismic layer 2 (largely volcanics and sheeted-dyke complexes) and seismic layer 3 (generally gabbroic and ultramafic units above Moho) (Fig. 4b) shows a striking resemblance to the faults and decoupling surfaces in the Agros window. Numerous planar faults cut seismic layer 2 (L-2), a sporadic set of antithetic faults cut the upper seismic layer 3 (L-3) and several sub-horizontal reflectors are observed discontinuously in layer 3. In DSDP/ODP hole 504B a fault zone has been localized within the pillow-dyke transition^{58,59} where ‘stock-work-like’ hydrothermal textures and assemblages suggest that fluids have been channelled along the fault. In the Josephine ophiolite, planar (domino) faults extend more than 1 km below the crustal section into a high-temperature (500–800 °C) horizontal shear zone^{18,19}. This high-temperature shear zone has been compared to a detachment fault of the ‘Basin-and-Range’ type, but domino-fault blocks overlying a sub-horizontal decoupling surface are common to both the Agros window and the Josephine ophiolite model. In the Agros window, however, the crush zone accommodates fault-block rotation above and below the dyke–pluton transition within the brittle regime and a lower-level, high-temperature decoupling zone is inferred to accommodate the rotation of the plutonics.

When viewed from the perspective of the Agros window we suggest that sub-horizontal reflectors in oceanic lithosphere could represent decoupling zones that have localized simultaneously at different lithospheric levels to generate stacked arrays. Depth-dependent strain-rate gradients, relief on lithological boundaries and the rheological contrasts across them would promote decoupling. The timing of displacements and displacement rates along these stacked decoupling zones would control their shear-sense histories. In this analogy, moderate-to-steeply dipping reflectors would be brittle faults that link into (or cut) progressively deeper sets of sub-horizontal decoupling surfaces as the lithosphere moves off-axis.

The polyphase shear zone and fault development indicated by our model would compartmentalize the lithosphere into lenses whose kinematic histories would be controlled by the depth-dependent displacement rates and geometries of the shear zones bounding them. If the structural geometries of our model for the Agros window were to dominate oceanic lithosphere deformation, the model would be broadly compatible with a regional-scale ‘pure-shear’ model. We emphasize, however, that we are proposing the Agros window model as an alternative for structures at the dyke–pluton boundary that could coexist with low-angle, listric normal detachments. Exposures of ultramafic rocks at ridge–transform intersections along the Mid-Atlantic Ridge have been attributed to listric normal detachment faulting, although their occurrence can be explained by alternative mechanisms and does not necessarily imply large (several kilometre) offsets along major shear zones^{46,61}. If we accept that both types of structures could exist, it is now important to evaluate their distribution at sea-floor spreading centres.

The mechanical contrasts generated by the magmatic construction of the oceanic lithosphere strongly influence the vertical partitioning of deformation, and could represent a fundamental difference between the rheology of oceanic and continental litho-

sphere. Alternatively, the Agros window may prove to be a viable analogue for continental deformation. □

Received 12 July 1994; accepted 16 February 1995.

- Mutter, J. C. & Karson, J. A. *Science* **257**, 627–634 (1992).
- Kong, L. S. L., Solomon, S. C. & Purdy, G. M. *J. geophys. Res.* **97**, 1659–1685 (1992).
- Smith, D. & Cann, J. R. *Nature* **365**, 707–715 (1993).
- Mutter, J. C. & North Atlantic Transect Study Group *Geology* **13**, 629–632 (1985).
- White, R. S. et al. *Geology* **18**, 462–465 (1990).
- Toomey, D. R., Purdy, G. M., Solomon, S. C. & Wilcock, W. S. D. *Nature* **347**, 639–645 (1990).
- Vera, E. E. et al. *J. geophys. Res.* **95**, 15529–15556 (1990).
- Sinton, J. M. & Detrick, R. S. *J. geophys. Res.* **97**, 187–216 (1992).
- Detrick, R. S. et al. *Science* **259**, 499–503 (1993).
- Morris, E. et al. *J. geophys. Res.* **98**, 13879–13903 (1993).
- McCarthy, J., Mutter, J. C., Morton, J., Sleep, N. & Thompson, G. *Bull. geol. Soc. Am.* **100**, 1423–1436 (1988).
- Macdonald, K. C. in *The Geology of North America* Vol. M (eds Vogt, P. R. & Tucholke, B. E.) 51–68 (Geol. Soc. Am., Boulder, 1986).
- Wernicke, B. P. & Burchfiel, B. C. *J. struct. Geol.* **4**, 105–115 (1982).
- Gibbs, A. D. *J. struct. Geol.* **5**, 153–160 (1983).
- Dick, H. J. B. et al. *Proc. ODP Sci. Res.* **118**, 359–398 (1991).
- Harper, G. D. *Tectonics* **93**, 4625–4656 (1985).
- Karson, J. A. et al. *Nature* **328**, 681–685 (1987).
- Norrell, G. T. & Harper, G. D. *Geology* **16**, 827–830 (1988).
- Norrell, G. T. & Harper, G. D. in *Ophiolites: Oceanic Crustal Analogues* (eds Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C.) 521–534 (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- Karson, J. A. in *Ophiolites: Oceanic Crustal Analogues* (eds Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C.) 547–555 (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- Moores, E. M. & Vine, F. J. *Phil. Trans. R. Soc. A268*, 443–466 (1971).
- Casey, J. F., Dewey, J. F., Fox, P. J., Karson, J. A. & Rosencrantz, E. in *The Oceanic Lithosphere* (ed. Emiliani, C.) 305–338 (Wiley-Interscience, New York, 1981).
- Gass, I. G. in *Ophiolites: Oceanic Crustal Analogues* (eds Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C.) 1–10 (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- Nicolas, A. *Structure of Ophiolites and Dynamics of Oceanic Lithosphere* (Kluwer Academic, London, 1989).
- Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C. (eds) *Ophiolites: Oceanic Crustal Analogues* (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- Miyashiro, A. *Earth planet. Sci. Lett.* **19**, 218–224 (1973).
- Varga, R. J. & Moores, E. M. *Geology* **13**, 846–850 (1985).
- Allerton, S. & Vine, F. J. *Geology* **15**, 593–597 (1987).
- Varga, R. J. *J. struct. Geol.* **13**, 517–537 (1991).
- Allerton, S. & Vine, F. J. *Geology* **19**, 637–640 (1991).
- Varga, R. J. & Moores, E. M. in *Ophiolites: Oceanic Crustal Analogues* (eds Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C.) 53–64 (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- Hurst, S. D., Moores, E. M. & Varga, R. J. *Tectonics* **13**, 139–156 (1994).
- MacLeod, C. J. in *Ophiolites: Oceanic Crustal Analogues* (eds Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C.) 75–86 (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- MacLeod, C. J. & Rothery, D. A. in *Ophiolites and their Modern Oceanic Analogues* (eds Parson, L. M., Murton, B. J. & Browning, P.) 39–63 (Geol. Soc., Lond., 1992).
- Baragar, W. R. A., Lambert, M. B., Baglow, N. & Gibson, I. in *Cyprus Crustal Study Project: Initial Reports, Hole CY-4* (eds Gibson, I., Malpas, J., Robinson, P. & Xenophontos, C.) 69–106 (Geol. Surv. Can., Ottawa, 1989).
- Allen, C. R. thesis, Univ. Cambridge (1975).
- Malpas, J., Case, G. & Moore, P. in *Cyprus Crustal Study Project: Initial Reports, Hole CY-4* (eds Gibson, I., Malpas, J., Robinson, P. & Xenophontos, C.) 31–37 (Geol. Surv. Can., Ottawa, 1989).
- Wilson, R. A. M. *The Geology of the Xeros-Troodos Area* (Mem. No. 1, Geol. Surv. Dept. Cyprus, Nicosia, 1959).
- Moores, E. M., Varga, R. J., Verosub, K. L. & Ramsden, T. in *Ophiolites: Oceanic Crustal Analogues* (eds Malpas, J., Moores, E. M., Panayiotou, A. & Xenophontos, C.) 27–35 (Geol. Surv. Dept. Cyprus, Nicosia, 1990).
- Robertson, A. & Xenophontos, C. in *Magmatic Processes and Plate Tectonics* (eds Prichard, H. M., Alabaster, T., Harris, N. B. W. & Neary, C. R.) 85–119 (Geol. Soc., Lond., 1993).
- Bear, L. M. *The Geology and Mineral Resources of the Agros—Akrotiri Area* (Mem. No. 2, Geol. Surv. Dept. Cyprus, Nicosia, 1960).
- Verosub, K. L. & Moores, E. M. *J. geophys. Res.* **86**, 6335–6349 (1981).
- Simonian, K. O. & Gass, I. G. *Bull. geol. Soc. Am.* **89**, 1220–1230 (1978).
- Murton, B. J. *J. geol. Soc. Lond.* **143**, 845–854 (1986).
- MacLeod, C. J., Allerton, S., Gass, I. G. & Xenophontos, C. *Nature* **348**, 717–720 (1990).
- Agar, S. M. *J. geophys. Res.* **99**, 3175–3200 (1994).
- Schiffman, P., Smith, B. M., Varga, R. J. & Moores, E. M. *Nature* **325**, 423–425 (1987).
- Delaney, J. R., Mogk, D. W. & Mottl, M. J. *J. geophys. Res.* **92**, 9175–9192 (1987).
- Harper, G. D., Bowman, J. R. & Kuhns, R. J. *J. geophys. Res.* **93**, 4625–4656 (1988).
- Bettison-Varga, L., Varga, R. J. & Schiffman, P. *Geology* **20**, 987–990 (1992).
- Vokes, F. M. in *UN Special Fund Project Report: Groundwater and Mineral Resources Survey, Cyprus* (Nicosia, 1965).
- Kidd, R. G. W. & Cann, J. R. *Earth planet. Sci. Lett.* **24**, 151–155 (1974).
- Karson, J. A., Hurst, S. D. & Lonsdale, P. *Geology* **20**, 685–688 (1992).
- Murton, B. J. *J. geol. Soc. Lond.* **143**, 845–854 (1986).
- Mevel, C. & Cannat, M. in *Ophiolite Genesis and Evolution of Oceanic Lithosphere* (eds Peters, T. J., Nicolas, A. & Coleman, R. G.) 293–312 (Kluwer Academic, London, 1991).
- Seyfried, W. E. Jr, Berndt, M. E. & Seewald, J. S. *Can. Mineralogist* **26**, 787–804 (1988).
- Auzende, J. M. et al. *Nature* **337**, 726–729 (1989).
- Agar, S. M. *J. Geodynamics* **13**, 119–140 (1991).
- Becker et al. *Proc. ODP Sci. Res.* **111**, 147–156 (1988).
- Alexander, R. J. & Harper, G. D. in *Ophiolites and their Modern Oceanic Analogues* (eds Parson, L. M., Murton, B. J. & Browning, P.) 3–38 (Geol. Soc., London, 1992).
- Cannat, M. et al. *Geology* **23**, 49–52 (1995).

ACKNOWLEDGEMENTS. This manuscript was significantly improved by informal discussion with J. Malpas, E. Moores, J. Cann, K. Devaux, G. Gee, K. Gillis, S. Hurst, J. Karson, H. Staudigel, R. Varga and C. Xenophontos, and formal reviews by S. Allerton, C. MacLeod, S. Lewis and J. Gardner. Graphics assistance was provided by J. Zwinakis and by means of the GMT system developed by P. Wessel and W. Smith. We are grateful to C. Xenophontos and the Hellenic Mining Company for logistical support on Cyprus. S.M.A. was supported by the US NSF.

LETTERS TO NATURE

Tidal effects of disconnected hydrocarbon seas on Titan

Stanley F. Dermott* & Carl Sagan†

* Department of Astronomy, PO Box 112055, University of Florida, Gainesville, Florida 32611-2055, USA

† Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853, USA

THERMODYNAMIC and photochemical arguments^{1–4} suggest that Titan, the largest satellite of Saturn, has a deep ocean of liquid hydrocarbons. At visible wavelengths, Titan's surface is obscured by a thick stratospheric haze, but radar observations^{5–7} have revealed large regions of high surface reflectivity that are inconsistent with a global hydrocarbon ocean. Titan's surface has also been imaged at infrared wavelengths^{8–10}, and the highest-resolution data (obtained by the Hubble Space Telescope) show clear variations in surface albedo and/or topography¹⁰. The natural interpretation of these observations is that Titan, like the Earth, has continents and oceans. But Titan's high orbital eccentricity poses a problem for this interpretation, as the effects of oceanic tidal friction would have circularized Titan's orbit for most configurations of oceans and continents^{1,11}. Here we argue that a more realistic topography, in which liquid hydrocarbons are confined to a number of disconnected seas or crater lakes, may satisfy both the dynamical and observational constraints.

Recent observations^{8–10} of the surface of Titan through windows in the infrared spectrum seem to confirm that the satellite is in synchronous rotation. In such a case, because the orbital eccentricity of Titan is high, Saturn raises two important tides on the satellite: a radial tide due to the variation of Titan's distance from Saturn and a librational tide due to the fact that Titan rotates at a uniform rate whereas its orbital motion is keplerian and non-uniform. Frictional damping of these tidal oscillations results in the circularization of Titan's eccentric orbit on a timescale determined by the energy dissipation rate or, equivalently, by Titan's tidal dissipation function, Q_T , which varies inversely with the dissipation rate. Assuming that tidal damping has operated over the age of the Solar System, the observation that Titan's eccentricity is still as high as 0.029 implies that Q_T is high (>200) and that the energy dissipation rate is low¹.

The Earth, which is the only body in the Solar System known to have an extensive ocean, has a tidal dissipation coefficient, Q_E , about 20 times smaller¹² than the lower limit on Q_T . This arises because land barriers modify the tidal flow and high rates of tidal dissipation occur in shallow seas and continental shelves (particularly the near-resonant shelves¹³) where the tidal currents are an order of magnitude greater than in the deep ocean^{12–14}. The existence of both near-global oceans and extensive land barriers on Earth accounts for the low value of Q_E and led us to argue¹ that Titan's ocean must either be deep and global, with no significant protruding land, or else insignificant. The recent radar and infrared observations of Titan's surface

seem to rule out a global ocean. We therefore need to reassess the nature and extent of Titan's hydrocarbon reservoirs.

The time-dependent part of the tide-raising potential of Saturn at some point on the surface of Titan with spherical polar coordinates (R_T, θ, ϕ) , where R_T is the radius of Titan and θ and ϕ are, respectively, the colatitude and longitude, is

$$V = -\frac{GM_S}{a} \left(\frac{R_T}{a} \right)^2 \times 3e \left[\frac{1}{2} (3 \sin^2 \theta \cos^2 \phi - 1) \cos nt + \sin^2 \theta \sin 2\phi \sin nt \right] \quad (1)$$

where G is the gravitational constant, M_S is the mass of Saturn, and a , n ($=2\pi/T$, where T is the orbital period of the satellite) and e ($=0.029$) are the semi-major axis, mean motion and eccentricity of Titan's orbit. If we neglect the inertia of the ocean and the effects of friction, then the variation with time t of the ocean depth produced by this potential is given by

$$\zeta(R_T, \theta, \phi; t) = 3eh \left[\frac{1}{2} (3 \sin^2 \theta \cos^2 \phi - 1) \cos nt + \sin^2 \theta \sin 2\phi \sin nt \right] \quad (2)$$

where h is the mean difference in the amplitudes of the oceanic and solid body tides and is estimated^{1,15} to be ~ 120 m. The first term in equation (2) is the radial tide; the second is the librational tide.

In estimating energy dissipation rates on Titan due to ocean tidal friction, an analogy with the Earth is probably inappropriate, for two separate reasons. First, tectonic forces on the Earth have shaped a planet with a two-thirds ocean cover and a small number of large, continental land masses with wide, gently-sloping shelves. In contrast, Titan's surface topography will have been determined by impacts similar to those that cratered the visible surfaces of its near neighbours and may have shattered some of the small, inner satellites of Saturn¹⁶. The solid surface of Titan may resemble that of its near neighbour Rhea in having a global distribution of comparatively steep-sided, unrelaxed basins and craters, although Titan's dense atmosphere may have protected its surface from impacts by small bodies¹⁷ resulting in a near-absence of craters of diameter $\lesssim 20$ km. The infrared images of Titan's surface taken by the Hubble Space Telescope¹⁰ do show a profusion of dark spots that could be associated with impact craters.

Second, resonance is important in tidal dissipation on the Earth, but this is almost certainly not the case on Titan. The periods, T_g , of the normal modes of a rectangular tidal basin are determined by the speed of the free gravity waves. From Merian's formula, we have^{14,18}

$$T_g = 2L/[N_m(gd)^{1/2}] \quad (3)$$

where L is the length (along the direction of wave propagation) and d the depth of the basin, g is the surface gravity and the mode number $N_m = 1, 2, 3, \dots$. Resonance arises when one of the periods of the normal modes equals the tidal period, or, in the case of Titan, the orbital period of the satellite, T . From equation (1), we obtain the resonant depth of a tidal basin, $d_{\text{res}} = 3L^2/(N_m^2 \pi G \rho R_T T^2)$, where ρ and R_T are the mean density and radius of the satellite. On Titan, for tidal periods equal to the orbital period, in the extreme case of $L \approx R_T$ and $N_m = 1$, $d_{\text{res}}/R_T \approx 4 \times 10^{-6}$ and $d_{\text{res}} \approx 10$ m. It is unrealistic to expect the solid surface of Titan to be topographically relaxed on scales of 10 m over lengths comparable with the satellite radius. Furthermore, because d_{res}/L decreases with decreasing L , smaller resonant tidal basins would have to be even shallower and smoother. Because the orbital period of Titan is long (15.95 days), the ocean's tidal oscillations are not near-resonant and the kinetic energy of the tidal flow is negligible in comparison with its potential energy. Tidal flow on Titan, whether the ocean is global or pooled in basins, is determined by the fact that all the fluid surfaces are in static equilibrium and bounded by equipotentials.

This fact allows simple methods to be used in determining the tidal dissipation rates.

The eccentricity damping timescale of Titan depends on the total rate of tidal dissipation and is determined by¹

$$\dot{e} = x \dot{E}_r / (e G M_S M_T) \quad (4)$$

where $\dot{E}_r$ is the rate of tidal dissipation due to the radial tide alone, M_T is the mass of Titan, and x is a factor that allows for the combined action of the librational and radial tides. Using a numerical hydrodynamical code, Sears¹¹ has determined that $x \approx 6$. If boundary-layer turbulence on the ocean floor is the dominant energy sink, then $\langle |\dot{E}_r| \rangle = 4\pi R_T^2 / \sigma \langle |v^3| \rangle$, where $f \approx 0.0042$ is the coefficient of skin friction^{11,19}, σ ($=0.65 \text{ g cm}^{-3}$) is the density of the putative hydrocarbon ocean, and v is the local velocity of the tidal current. Assuming that the ocean is uniform and global, and averaging over the surface of the satellite and over one tidal cycle, we obtain

$$\langle |v^3| \rangle = 2\gamma \quad (5)$$

where $\gamma = (3ehnR_T/d)^3 / (105\pi^2)$. Thus, the mean velocity of the tidal current and the dissipation rate decrease as the depth of the ocean increases. Given that the initial eccentricity of Titan must be less than unity, integration of equation (4) over the age, t_o , of the ocean places an upper bound on $\dot{E}_r$ and hence a lower bound on the minimum depth, d_{min} , of the ocean. d_{min} varies as $t_o^{1/3}$ and is thus only weakly dependent on t_o . By assuming that the ocean is as old as the Solar System, we obtain $d_{\text{min}} \approx 0.6$ km. (Our previous estimation¹ of a lower limit on d_{min} of 0.4 km omitted a factor of 2 in the equation relating $\dot{E}_r$ to $\dot{e}$).

Current thermodynamic and photochemical depth estimates⁴ of Titan's hydrocarbon ocean are in the range 0.3–0.6 km, but may be less because of the uncertain effects of impact chemistry, cosmic rays and clathration. Thus, if the tidal currents are near-global, then the flow and dissipation rates on Titan seem to be high and the eccentricity damping timescale much less than the age of the Solar System. This conclusion is supported by the numerical hydrodynamical calculations of Sears¹¹. A cratered topography may provide a solution to this dilemma²⁰. If the fluid is pooled in basins and thus confined to a number of disconnected, land-locked seas and crater lakes, then the dissipation rate may be substantially reduced. To quantify this reduction, we analyse a simple analogy for the tidal flow associated with the radial tide. Consider a square tray with sides εR_T (where ε is a constant roughly equal to 1) filled with fluid to a depth d and oscillating harmonically with an amplitude $3eh$ and frequency n about an axis through its centre and parallel to one of its sides. If the tray is shallow, but $d \gg d_{\text{res}}$, then $|v^3|$, averaged over the surface of the tray and over one cycle, is given by $\langle |v^3| \rangle = \pi \varepsilon^3 \gamma$. The similarity of this expression to equation (5) shows that we have captured the essential physics of the problem. If the tray is now divided (like an ice-tray) into N_c similar cells, where $\sqrt{N_c} = (\varepsilon R_T/L)$ where L is the length of each cell) is an integer, then the mean dissipation rate is reduced by a factor $N_c^3 = (\varepsilon R_T/L)^6$. This is because the distance over which the fluid flows and the amplitude of oscillation in each cell are both reduced by a factor $\sqrt{N_c}$, reducing the local velocity v by a factor N_c .

This simple model suggests that confining the fluid on Titan to a number of disconnected seas or crater lakes greatly extends the eccentricity damping timescale, Γ_e . If the ocean is global and has a depth $d \geq d_{\text{min}} = 0.6$ km, then $\Gamma_e \geq t_{\text{ss}}$, the age of the Solar System. But if the fluid is confined to N_c basins having depths over about d_{min} and a total area comparable with that of the satellite, then $\Gamma_e \geq N_c^3 t_{\text{ss}}$. The eccentricity damping timescale decreases with d^3 and could be $\sim t_{\text{ss}}$ if the depths of the basins are less than d_{min} by the same factor N_c , but this requires the basins to be unrealistically shallow and smooth. For example, the case $\Gamma_e \geq t_{\text{ss}}$ and $N_c \approx 100$ requires basins of diameters $\geq \varepsilon R_T / \sqrt{N_c} \geq 300$ km to be topographically relaxed on scales $\lesssim d_{\text{min}}/N_c \approx 6$ m.

If we consider a realistic, spherical model of Titan's surface and allow that the craters or tidal basins have a spectrum of sizes and that their depths vary with their diameters, then the dissipation rate is determined by the frequency distribution of crater sizes. Observations²¹ of the other satellites of Saturn show that this is well represented by the power law

$$dN = AD^b dD \quad (6)$$

where dN is the number of craters in the diameter range D to $D + dD$ and $b = -3$. The constant A in equation (6) is determined by the normalization requirement that the total area of the craters be comparable with that of the satellite. If the craters range in diameter from $D_{\max}$ to $D_{\min}$, then $A = 4\pi R_T^2 / \ln(D_{\max}/D_{\min})$ and, if $D_{\max} \approx R_T$ and $D_{\min} \approx 20$ km, then $\ln(D_{\max}/D_{\min}) \approx 5$. If we compare the dissipation rate in a global ocean of depth d_o with that of a cratered satellite on which all the craters have the same depth d_o , then the dissipation rate is reduced by a factor $\sim 6 (1.32R_T/D_{\max})^6 \ln(D_{\max}/D_{\min})$. If the depth d of each crater decreases linearly with its diameter according to $d = (D/D_{\max})d_o$, then the dissipation rate is reduced by a factor $\sim 3 (1.32R_T/D_{\max})^6 \ln(D_{\max}/D_{\min})$. In both cases, the dissipation rate is an average rate but is chiefly determined by the dissipation that occurs in the largest tidal basin. Thus, on Titan, the actual dissipation rate will vary with the location of that largest crater.

This model demonstrates that if the fluid on Titan is confined to shallow seas with sizes smaller than the satellite radius, then as long as the seas are land-locked and disconnected, large liquid hydrocarbon reservoirs, and regions with no liquid hydrocarbons can coexist on the surface of Titan along with a large orbital eccentricity. The recent Hubble Space Telescope observations¹⁰ of Titan show variations in the surface reflectivity of 10–20%. The interpretation of such variations may provide a test of the configuration of land and sea proposed here. The configuration could also be tested by the Arecibo Observatory after its current resurfacing, and by radar and near-infrared observations by the Cassini Saturn Orbiter.

We have assumed that the largest source of tidal dissipation in Titan's ocean is boundary-layer turbulence. Our arguments also apply, however, to any other source of oceanic dissipation that increases with the velocity of the tidal current, for example, dissipation associated with internal waves¹³. Finally, we note that even if oceanic tidal friction is negligible on Titan this may not entirely solve the problem of Titan's high orbital eccentricity because solid body friction may reduce Q_T to a value less than 200 (ref. 22). One possible, but unlikely, solution is that Titan's eccentricity is not primordial but is the result of a recent large, but less than cataclysmic, collision (one less than 10^9 years ago). But even in the most favourable circumstances, this would require Titan to have been struck by, for example, a body with a mass $\sim 0.005M_T$ (or diameter $\sim 10^3$ km) moving with a velocity ~ 15 km s⁻¹ relative to Saturn. No impact basin of appropriate size seems evident in the current infrared data for Titan. □

16. Smith, B. A. *et al.* *Science* **212**, 163–191 (1981).
17. Engel, S., Hartmann, W. K. & Lunine, J. I. (abstr.) *Bull. Am. astr. Soc.* **26**, 1183 (1994).
18. Merian, J. R. *Über die Bewegung tropfbarer Flüssigkeiten in Gefässen* (Basle, 1828).
19. Turcotte, D. L. & Schubert, G. *Geodynamics* (Wiley, New York, 1982).
20. Sagan, C. & Dermott, S. (abstr.) *Bull. Am. astr. Soc.* **26**, 1183 (1994).
21. Chapman, C. R. & McKinnon, W. B. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 492–580 (Univ. Arizona Press, Tucson, 1986).
22. Peale, S. J., Cassen, P. & Reynolds, R. T. *Icarus* **43**, 65–72 (1980).

ACKNOWLEDGEMENTS. We thank J. Lunine for comments. This work was supported by NASA.

Comb-type grafted hydrogels with rapid de-swelling response to temperature changes

Ryo Yoshida, Katsumi Uchida*, Yuzo Kaneko*, Kiyotaka Sakai*, Akihiko Kikuchi, Yasuhisa Sakurai & Teruo Okano†

Institute of Biomedical Engineering,
Tokyo Women's Medical College, 8-1 Kawada-cho, Shinjuku-ku,
Tokyo 162, Japan

* Department of Chemical Engineering, Waseda University,
3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169, Japan

MANY polymeric hydrogels undergo abrupt changes in volume in response to external stimuli such as changes in solvent composition¹, pH², electric field³ and temperature^{4–6}. For several of the potential applications of these materials, such as 'smart' actuators, a fast response is needed. The kinetics of swelling and de-swelling in these gels are typically governed by diffusion-limited transport of the polymeric components of the network in water, the rate of which is inversely proportional to the square of the smallest dimension of the gel^{7–9}. Several strategies have been explored for increasing the response dynamics^{10–14}, such as introducing porosity¹⁴. Here we show that we can induce rapid de-swelling of a polymer hydrogel by tailoring the gel architecture at the molecular level. We prepare a crosslinked hydrogel in which the polymer chains bear grafted side chains; the latter create hydrophobic regions, aiding the expulsion of water from the network during collapse. Whereas similar gels lacking the grafted side chains can take more than a month to undergo full de-swelling, our materials collapse in about 20 minutes.

Poly(*N*-isopropylacrylamide) (PIPAAm) is soluble in aqueous media at solution temperatures below 32 °C, its critical solution temperature¹⁵. Above this point, it undergoes a discontinuous phase transition, precipitating from solution suddenly and reversibly over a narrow temperature range^{1,16,17}. In contrast to conventional crosslinked PIPAAm hydrogels, we have prepared a thermosensitive hydrogel with a comb structure in which PIPAAm chains are grafted onto crosslinked networks (Fig. 1). Within the gel, terminally grafted chains have freely mobile ends, distinct from the typical network structure in which both ends of the PIPAAm chains are crosslinked and relatively immobile. With increasing temperature, grafted PIPAAm chains begin to collapse from their expanded (hydrated) form to compact (dehydrated) forms. This collapse occurs before the PIPAAm network begins to shrink, because of the mobility of the grafted chains. The grafted polymer chains dehydrate to create hydrophobic nuclei which enhance aggregation of the crosslinked chains.

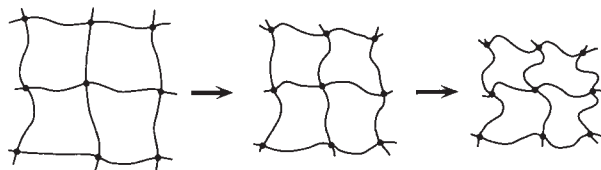
The transmittance of PIPAAm aqueous solution changes rapidly because of conformational changes over a small temperature change^{15,17}. We have previously observed changes from hydro-

Received 3 November 1994; accepted 6 February 1995.

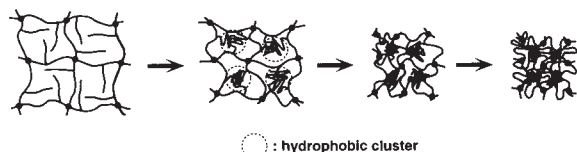
1. Sagan, C. & Dermott, S. F. *Nature* **300**, 731–733 (1982).
2. Fiasar, F. M. *Science* **221**, 55–57 (1983).
3. Lunine, J. I., Stevenson, D. J. & Yung, Y. K. *Science* **222**, 1229–1230 (1983).
4. Lunine, J. I. *Rev. Geophys.* **31**, 133–149 (1993).
5. Muhleman, D. O., Grossman, A. W., Butler, B. J. & Slade, M. A. *Science* **248**, 975–980 (1990).
6. Muhleman, D. O., Grossman, A. W., Slade, M. A. & Butler, B. J. (abstr.) *Bull. Am. astr. Soc.* **24**, 954–955 (1992).
7. Grossman, A. W. & Muhleman, D. O. (abstr.) *Bull. Am. astr. Soc.* **24**, 954 (1992).
8. Lemmon, M. T., Karkoschka, E. & Tomasko, M. *Icarus* **103**, 329–332 (1993).
9. Griffith, C. A. *Nature* **364**, 511–514 (1993).
10. Smith, P. H., Lemmon, M. T., Caldwell, J. J., Allison, M. D. & Sromovsky, L. A. (abstr.) *Bull. Am. astr. Soc.* **26**, 1181 (1994).
11. Sears, W. D. (abstr.) *Bull. Am. astr. Soc.* **26**, 1184 (1994); *Icarus* (in the press).
12. Lambeck, K. *The Earth's Variable Rotation: Geophysical Causes and Consequences* 295 (Cambridge Univ. Press, Cambridge, 1980).
13. Webb, D. J. *Contemp. Phys.* **23**, 419–442 (1982).
14. Burns, J. A. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 117–158 (Univ. Arizona Press, Tucson, 1986).
15. Dermott, S. F. *Icarus* **37**, 310–321 (1979).

† To whom correspondence should be addressed.

Homopolymer gel



Comb-type graft polymer gel



○ : hydrophobic cluster

philic to hydrophobic behaviour for PIPAAm-terminally grafted surfaces due to rapid changes in polymer conformation (again, this is attributed to the mobility of grafted PIPAAm¹⁸). PIPAAm-grafted surfaces show hydrophilic properties at low temperatures; as the temperature increases, the contact angle of water also increases, indicating increasingly hydrophobic surface properties. The behaviour of these systems is distinct from that of surfaces on which PIPAAm molecules were attached by anchoring at several points along the chain: the decrease in hydrophilicity with increasing temperature is substantially less for the latter than for surfaces with terminally grafted chains.

Above the phase transition temperature, PIPAAm networks undergo both inter- and intramolecular hydrophobic interactions between alkyl side groups, which induce polymer network aggregation. We have reported previously¹⁹ that hydrophobic butyl methacrylate as a comonomer increases the intrinsic aggregation forces of PIPAAm gels. Typically, a dense collapsed polymer layer rapidly forms on the surface of these hydrophobized gels above their phase transition temperature¹⁹.

The hydrated and expanded grafted PIPAAm chains of our new gels are hydrophilic at low temperature. With increasing temperature, the grafted PIPAAm rapidly starts to dehydrate and becomes more hydrophobic. This thermally induced dehydration of the grafted PIPAAm chains promotes rapid shrinking of the crosslinked network.

We prepared comb-type PIPAAm grafted gels as follows: first, we synthesized semitelechelic (that is, having one functional end group per polymer chain) PIPAAm with a terminal amino end group by radical telomerization of the monomer (IPAAm) with 2-aminoethanethiol (AESH) as a chain transfer agent. We dissolved 13.4 g of purified IPAAm, 0.128 g of AESH and 0.0197 g of *N,N'*-azobisisobutyronitrile as an initiator in 50 ml of distilled *N,N*-dimethylformamide (DMF). The ampoule containing the solution was sealed by conventional methods and immersed in a water bath held at 75 °C for 15 h. After concentrating the

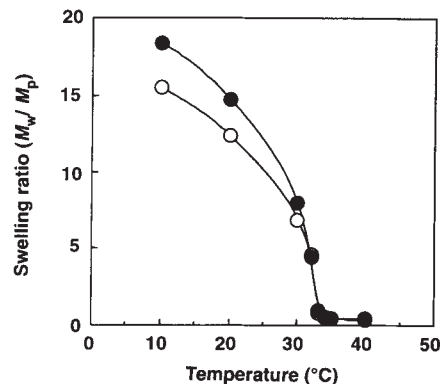
FIG. 1 Structure and shrinking mechanisms for conventional homopolymer and comb-type grafted PIPAAm gels undergoing temperature-induced collapse in aqueous media.

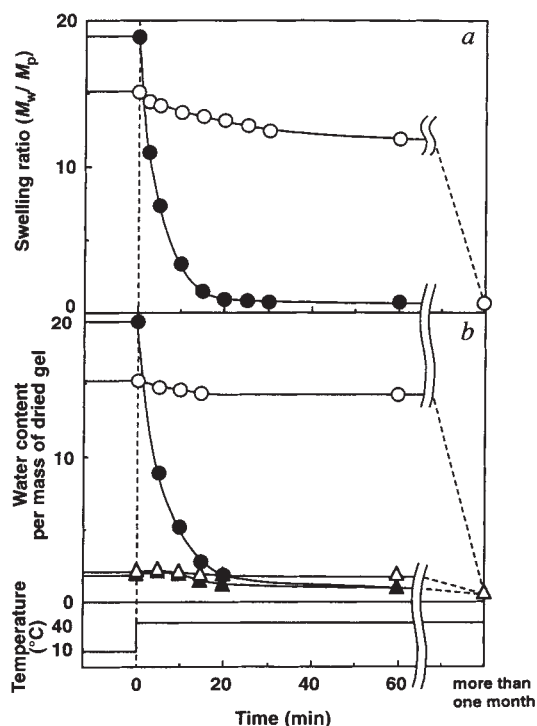
reactant by DMF evaporation, the reactant was poured into diethyl ether to precipitate the synthesized semitelechelic PIPAAm. The PIPAAm product was collected over a filter and purified by repeated precipitation in diethyl ether. The number-average and mass-average relative molecular masses determined by gel permeation chromatography in tetrahydrofuran at 40 °C (against polystyrene standards) were 14,000 and 19,000, respectively. A polymerizable end group was introduced into the semitelechelic amino PIPAAm by reacting 2 g of the amino-semitelechelic PIPAAm with 0.4 g of *N*-acryloxysuccinimide in 30 ml of DMF at 4 °C for 2 days.

To synthesize comb-type PIPAAm gels, we dissolved 1.09 g of IPAAm monomer, 0.468 g of semitelechelic PIPAAm containing a terminal acrylate group, 0.0266 g of *N,N'*-methylenebisacrylamide as a crosslinker and 48 µl of tetramethylethylenediamine as an accelerator in 10 ml of distilled water, and bubbled the solution with dry nitrogen gas for 10 minutes. Ammonium persulphate (8 mg, initiator) was added to the solution, and then the solution was injected between two Mylar sheets separated by a Teflon gasket (2.0 mm thick) and backed by glass plates. The solution was polymerized at 15 °C for 1 day. We immersed the gel membranes formed in pure water to remove unreacted compounds for 5 days, changing the water every day. The swollen gel membrane was cut into disks (15 mm diameter) using a cork borer and dried under ambient conditions for 1 day and under vacuum for 5 days at room temperature. The prepared grafted PIPAAm hydrogel is transparent, suggesting that the network structure is homogeneous on lengthscales above visible wavelengths and that the PIPAAm macromonomer radicals are active enough to copolymerize readily with small IPAAm monomers.

Measurement of equilibrium swelling ratios for the gels in water over a series of temperatures shows that the comb-type PIPAAm graft gel has the same phase transition temperature (32 °C) as the IPAAm homopolymer gel containing the same

FIG. 2 Equilibrium swelling ratio of homopolymer (○) and comb-type grafted PIPAAm (●) gels in water as a function of temperature. The swelling ratio is defined as the mass of absorbed water per mass of dried polymer disk (M_w/M_p).





amount of IPAAm (Fig. 2). Below this transition temperature, the comb-type graft gel shows slightly more swelling than the homopolymer gel. We speculate that the comb-type gel becomes more hydrated because of the inherent mobile nature of grafted PIPAAm chains, which have free ends that are readily exposed to water.

Disks of comb-type graft PIPAAm gels and IPAAm homopolymer gels with identical thicknesses and diameters show different de-swelling kinetics when the temperature is increased from below to above the phase transition temperature (Fig. 3a). Conventional IPAAm homopolymer gels shrink very slowly after the temperature is increased from 10 °C to 40 °C, requiring more than a month to reach equilibrium. This gel shrinks gradually from the surface inwards, mediated by diffusion of the collapsing polymer network and the release of entrapped water from the collapsing gel. As a dense, collapsed polymer layer impermeable to water is formed near the gel surface before bulk gel collapse is initiated (thermal convection is more rapid than mass transfer), gel shrinking is hindered after the initial stages by internal hydrostatic pressure^{20,21}. As a result, the PIPAAm homopolymer gel shrinkage rate is limited by water permeation from the gel interior through the collapsed polymer skin, keeping water within the gel for longer periods.

In contrast to the IPAAm homopolymer gel, the comb-type PIPAAm graft gel shrinks rapidly to its equilibrium state. In the process, the gel undergoes large, rapid volume changes with marked mechanical buckling, indicative of the much greater aggregation forces operating within the grafted gel. Trapped water is rapidly squeezed out from the gel interior. This is sup-

FIG. 3 a, Shrinking kinetics for homopolymer (○) and comb-type grafted PIPAAm (●) gels. The disk-shaped IPAAm homopolymer gel and the comb-type graft polymer gel (diameter 15 mm, thickness 2 mm) were first equilibrated in pure water at 10 °C. Gels were then quickly moved into water at 40 °C (the heating history is shown at the bottom of the figure). At specific times these gels were removed from the water and weighed after being wiped with filter paper to remove excess water on the surface. The shrinking kinetics are defined as the changes with time in the swelling ratios for the gels. b, Changes in water structure for shrinking homopolymer (○, freezable water; △, non-freezable water) and comb-type grafted PIPAAm (●, freezable water; ▲, non-freezable water) gels at 40 °C. Two types of gel disks equilibrated at 10 °C (diameter 15 mm, thickness 2 mm) in pure water were quickly transferred into water at 40 °C. The gels were cut into disks (4 mm diameter at centre) at predetermined times using a cork borer. After wiping excess water from the surface, the samples were sealed in aluminium pans and placed in a differential scanning calorimeter (Mettler TA-3000 system). After cooling to -20 °C within 5 min, the samples were heated to 40 °C at the rate of 1 °C min⁻¹. The endothermal heat of water in the gels was measured using an empty aluminium pan as a reference. From the integral of the endothermal peak for water melting (near 0 °C), the gel freezable water content was determined. Total water content was determined from the mass difference between the swollen gel and the dried gel. Non-freezable water content, restricted in its molecular mobility by interaction with hydrated polymer chains, was calculated from the difference between total and freezable water content for these gels.

ported by the changes in the amount of freezable water within two types of gel matrices determined by DSC. As can be seen in Fig. 3b, about 90% of the freezable water was desorbed from the graft-type PIPAAm gel within 20 min after the temperature was increased to 40 °C, although non-freezable water content remains nearly constant and fairly low. Therefore, we attribute the marked swelling changes observed for graft-type PIPAAm gel to the large change in freezable water content in the gel. In contrast to the graft-type PIPAAm gel, only 5% of the freezable water was desorbed from the IPAAm homopolymer gel after 60 min. Rapid shrinking of the graft-type PIPAAm gel is due to the immediate dehydration of PIPAAm grafted chains in the gel matrix followed by subsequent hydrophobic interactions between dehydrated grafted chains preceding shrinkage of the PIPAAm network, affecting rapid expulsion of water from the gel matrix. The rapid shrinking and release of most of the freezable water results in a large volume change, which suggests that a skin structure that would reduce the de-swelling rate is not formed in this case. An increase in void volume within the grafted PIPAAm network resulting from collapsed grafted chains may also contribute to rapid release of water to the gel exterior.

In the case of a gel swelling in water at 10 °C from an equilibrium shrunken state (40 °C), the swelling rate was slower than the de-swelling rate, and no difference in swelling rate was observed between the two types of gel. In both cases the amount of absorbed water increases in proportion to the square root of time. These results indicate that polymer network diffusion is rate-determining for swelling. The details of this process are currently under investigation. □

Received 30 June 1994; accepted 6 February 1995.

- Hirokawa, E. & Tanaka, T. *J. chem. Phys.* **81**, 6379–6380 (1984).
- Tanaka, T. *et al. Phys. Rev. Lett.* **45**, 1636–1639 (1980).
- Tanaka, T., Nishio, I., Sun, S.-T. & Ueno-Nishio, S. *Science* **218**, 467–469 (1981).
- Okano, T., Yui, N., Yokoyama, M. & Yoshida, R. in *Japanese Technology Reviews Section E* Vol. 4 (eds Ikoma, T. *et al.*) 67–105 (Gordon Science, Yverdon, Switzerland, 1993).
- Hoffman, A. S. *J. Controlled Release* **6**, 297–305 (1987).
- Dusek, K. (ed.) *Responsive Gels: Volume Transitions II* (Springer, Berlin, 1993).
- Tanaka, T. & Fillmore, D. J. *J. chem. Phys.* **70**, 1214–1218 (1979).
- Tanaka, T., Sato, E., Hirokawa, Y., Hirotsu, S. & Peetermans, J. *Phys. Rev. Lett.* **55**, 2455–2458 (1985).
- Sato Matsuo, E. & Tanaka, T. *J. chem. Phys.* **89**, 1695–1703 (1988).
- Hirasa, O., Ito, S., Yamauchi, A., Fujishige, S. & Ichijo, H. in *Polymer Gels* (eds DeRossi, D. *et al.*) 247–256 (Plenum, New York, 1991).

- Suzuki, M. in *Polymer Gels* (eds De Rossi, D. *et al.*) 221–236 (Plenum, New York, 1991).
- Kabra, B. G. & Gehrke, S. H. *Polym. Commun.* **32**, 322–323 (1991).
- Wu, X. S., Hoffman, A. S. & Yager, P. J. *Polym. Sci., A. Polym. Chem.* **30**, 2121–2129 (1992).
- Dong, L. C. & Hoffman, A. S. *J. Controlled Release* **13**, 21–31 (1990).
- Heskins, M., Guillet, J. E. & James, E. J. *Macromol. Sci. Chem.* **A2**, 1441–1455 (1968).
- Bae, Y. H., Okano, T. & Kim, S. W. *J. Polym. Sci. Polym. Phys.* **28**, 923–936 (1990).
- Takei, Y. G. *et al. Bioconj. Chem.* **4**, 341–346 (1993).
- Takei, Y. G. *et al. Macromolecules* **27**, 6163–6166 (1994).
- Yoshida, R., Sakai, K., Okano, T. & Sakurai, Y. *J. Biomater. Sci. Polym. Edn* **6**, 585–598 (1994).
- Yoshida, R. *et al. J. Biomater. Sci. Polym. Edn* **3**, 155–162 (1991).
- Yoshida, R., Sakai, K., Okano, T. & Sakurai, Y. *J. Biomater. Sci. Polym. Edn* **3**, 243–252 (1992).

ACKNOWLEDGEMENTS. We thank D. W. Grainger for comments and discussions.

Transformation of stishovite to a denser phase at lower-mantle pressures

Kathleen J. Kingma, Ronald E. Cohen,
Russell J. Hemley & Ho-kwang Mao

Geophysical Laboratory and Center for High-Pressure Research,
Carnegie Institution of Washington, 5251 Broad Branch Road, N.W.,
Washington DC 20015, USA

WHETHER stishovite, the highest-pressure polymorph of SiO_2 known from natural samples, transforms to a denser structure at higher pressures has long been of interest. A suggested transition from rutile to the CaCl_2 structure driven by a vibrational-mode instability¹ was supported by the observation of a frequency decrease (softening) of a Raman mode with increasing pressure². Subsequent X-ray diffraction measurements provided evidence³ for stability of the CaCl_2 phase near 100 GPa. Electronic-structure calculations predict, however, that the transition occurs at much lower pressure, where a shear modulus vanishes and before the Raman mode softens completely⁴. Here we use *in situ* Raman spectroscopy and a new theoretical model to investigate the high-pressure behaviour of stishovite. At 50 GPa, the pressure dependence of the soft B_{1g} mode abruptly changes and the E_g mode splits, as predicted for transformation to the CaCl_2 structure. Our results demonstrate that any free silica in the deep mantle (below 1,200–1,500 km) will exist in the CaCl_2 structure at considerably lower pressures than previously thought³.

Study of stishovite under mantle conditions has been difficult. The bulk modulus of stishovite, although now well constrained from single-crystal Brillouin scattering⁵, has been controversial, with values varying significantly between static compression diffraction studies^{3,6–9}. These discrepancies arise because compression of stishovite is highly anisotropic and sensitive to non-hydrostatic stress conditions. Although *in situ* X-ray diffraction data suggested that silica was stable in the CaCl_2 structure above ~100 GPa (ref. 3) the nonhydrostatic conditions of these experiments precluded the unambiguous determination of the transition pressure. Because the atomic displacements from the rutile to the CaCl_2 structure are subtle, and the volume change at the transition is likely to be small⁴, diffraction methods may not detect the onset and initial stages of this transformation. The vibrational spectrum is an important diagnostic of mineral structure and is complementary to diffraction techniques¹⁰. Spectroscopic measurements for stishovite^{2,11–15} date back to the mineral's discovery^{16,17}. Controversy surrounded the Raman spectrum of stishovite because of its very low scattering cross-section and problems with sample purity (see ref. 14). This was resolved with the development of new techniques^{2,14}, advances which also permitted measurement of the Raman spectrum to 32 GPa. The observed softening of the B_{1g} mode suggested² a transition below 100 GPa.

A number of theoretical calculations of post-stishovite phases have been performed, including simple ionic models, molecular-cluster methods, and crystalline electronic-structure computations (see ref. 18 for review). Ionic-model calculations have predicted a range of rutile-to- CaCl_2 transition pressures for silica, mostly above 100 GPa. Recently, however, calculations using the linear-augmented plane-wave (LAPW) method predicted the transition should occur at 45 GPa where the C_{11} – C_{12} elastic shear modulus vanishes⁴. These calculations employed the local density approximation (LDA) for the quantum-mechanical exchange and correlation interactions; no experimental data were used. According to the LAPW results, the soft rutile B_{1g} mode should become a hard mode of A_g symmetry at the transition. This shows that the rutile-to- CaCl_2 transition in silica

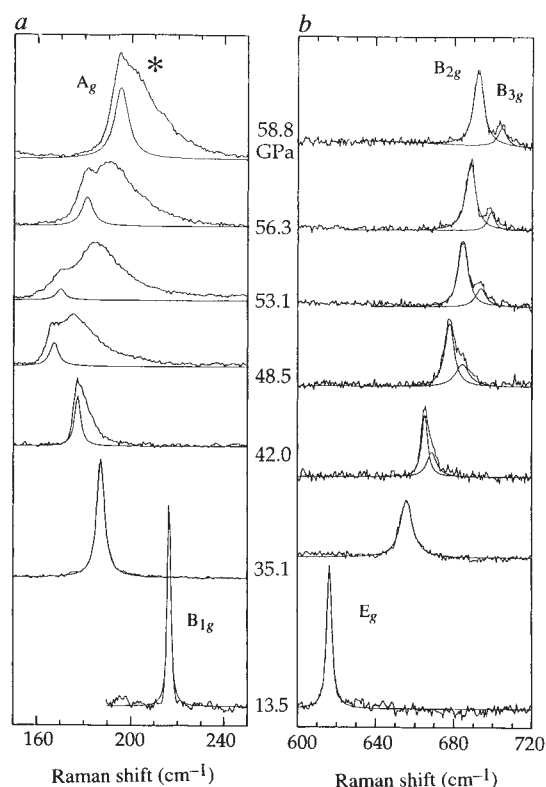


FIG. 1 Representative Raman spectra showing evidence for the rutile-to- CaCl_2 transition in silica. Lorentzian fits to the bands predicted by theory are shown as smooth traces. *a*, Initial frequency decrease (softening) of the rutile B_{1g} mode with pressure followed by frequency increase (hardening) of the CaCl_2 A_g mode with which the rutile B_{1g} is correlated. A broad band which is not predicted by theory was observed above 40 GPa (marked by asterisk in topmost panel). The new broad band is highly dependent on stress conditions; in our nonhydrostatic experiments, a broad feature was observed to split from the rutile B_{1g} band at ~20 GPa. Similar asymmetric broadening of the B_{1g} fundamental was noticed at the highest pressures of the previous experiment² which used the argon pressure medium (less hydrostatic than neon). This may arise from stress at grain-grain interfaces or at boundaries of twin domains which form at the transition, or from a further reduction in symmetry (possibly associated with residual stress inhomogeneities). Low-frequency, broad bands associated with multiphonon scattering are also observed²⁹ in α -quartz-structured SiO_2 and AlPO_4 . *b*, Splitting of the rutile E_g mode into the CaCl_2 B_{2g} and B_{3g} modes. All measurements were made at room temperature (298 K) using a Dilor XY spectrometer equipped with a modified optical microscope, custom-built confocal imaging optics, and a liquid-nitrogen-cooled CCD (charge-coupled device) detector. An argon-ion laser was used as the excitation source; 488.0- and 514.5-nm lines were used to check for Raman activity of observed bands. The laser radiation was focused to sample areas <5 μm in diameter. Pressures were determined with the ruby-fluorescence method (quasi-hydrostatic scale)³⁰. With neon as the medium, pressure gradients never exceeded 0.5 GPa.

should have a characteristic, and indeed unusual, signature in the high-pressure Raman spectrum. Motivated by the excellent agreement between the calculated⁴ and experimentally observed² mode behaviour, we have investigated the vibrational properties of stishovite using new Raman scattering techniques to test this theoretical prediction. Factor-group analysis also shows that the rutile E_g should split into B_{2g} and B_{3g} modes at the transition.

Experiments were performed using a Mao–Bell diamond-anvil cell¹⁹. Type I diamonds with very low fluorescence were used to limit the background signal. Starting materials for all experiments were synthesized in a multi-anvil press from natural

TABLE 1 Frequencies and pressure shifts of observed Raman-active bands*

Stishovite (1 bar)			CaCl ₂ -structured SiO ₂ (50 GPa)		
Symmetry†	ν_i (cm ⁻¹)	$(d\nu_i/dP)_T$ (cm ⁻¹ GPa ⁻¹)	Symmetry†	ν_i (cm ⁻¹)	$(d\nu_i/dP)_T$ (cm ⁻¹ GPa ⁻¹)
B _{2g}	966.2 (5)	3.68 (6)	B _{1g}	(Not observed)	
A _{1g}	754.1 (3)	3.12 (7)	A _g	906	2.2 (2)
E _g	589 (1)	2.09 (2)	B _{2g}	684	0.99 (8)
			B _{3g}	706	2.1 (2)
			Broad band	177	3.3 (2)
B _{1g}	231.6 (3)	-0.93 (2)	A _g	163	3.2 (2)

* Frequencies (ν_i), and isothermal pressure shifts ($d\nu_i/dP$) for the observed Raman-active bands of stishovite at 1 bar and CaCl₂-structured SiO₂ at 50 GPa (298 K).

† Irreducible representations for the zone-centre optical modes of vibration, Γ_{op} , are calculated for the rutile and CaCl₂ structures by factor-group analysis. Rutile ($P4_2/mnm, D_{4h}^{14}$), $\Gamma_{op} = A_{1g}^R + A_{2g}^R + A_{2u}^R + B_{1g}^R + B_{2g}^R + 2B_{1u}^R + E_g^R + 3E_u^R$; CaCl₂ ($Pnmm, D_{2h}^{12}$), $\Gamma_{op} = 2A_g^R + 2A_u + 2B_{1g}^R + B_{2g}^R + B_{3g}^R + B_{1u}^R + 3B_{2u}^R + 3B_{3u}^R$. Here superscript 'R' denotes Raman-active modes, superscript 'IR' denotes infrared-active modes, and the unmarked modes are neither Raman- nor infrared-active. The symbols A, B and E indicate the vibrational symmetry, where A and B denote one-dimensional, and E two-dimensional modes. Subscript g indicates vibrations which are symmetric with respect to a centre of inversion, and subscript u indicates those which are antisymmetric with respect to inversion.

‡ In correlating modes, $a_{1(rutile)} = a_{1(CaCl_2)}$, $a_{2(rutile)} = b_{1(CaCl_2)}$ and $c_{1(rutile)} = c_{1(CaCl_2)}$, where a, b and c are lattice parameters.

quartz. Care was taken to avoid contamination that complicated other studies^{13,14}. Because hydrostatic conditions were essential, stishovite grains (1–5 μ m) were compressed using neon as the pressure-transmitting medium; neon remains nearly hydrostatic (quasi-hydrostatic) to significantly higher pressure²⁰ than media used in previous work (for example, no medium³, salt¹⁵ and argon²). Stishovite powder was also compressed without a medium to examine nonhydrostatic effects.

During quasi-hydrostatic compression of stishovite, each of the four Raman-active modes predicted for the rutile structure were observed; pressure shifts (Table 1) are consistent with results of the earlier lower-pressure study². At 40 GPa, the stishovite B_{1g} mode showed significant asymmetry, and by 50 GPa two peaks could be resolved (Fig. 1)—a sharp peak that lies along the trend of the negative pressure shift of the stishovite B_{1g} and a broad band centred ~ 15 cm⁻¹ higher in frequency than the sharp B_{1g}. With further compression, the pressure shift of both bands suddenly changed, becoming positive (Fig. 2). The stishovite E_g mode was asymmetric by 40 GPa, and two peaks were clearly resolved above 50 GPa (Fig. 1). A slight break in the slope of the pressure shift of the strong stishovite A_{1g} mode

was observed at 50 GPa (Fig. 3), and the pressure dependence of the weak high-frequency stishovite B_{2g}, which was not observed previously under compression², could be determined (Table 1). This band was lost above 48 GPa because of its weakness and the increase in background signal with increasing pressure. All changes were reversible, and no hysteresis could be detected on decompression (Fig. 2).

The abrupt change in the pressure-dependence shift of the lowest-frequency mode occurs at a pressure of 50(± 3) GPa; the uncertainty arises from fitting the overlapping low-frequency peaks (Fig. 1). The Raman results reported here are in remarkable agreement with the earlier LAPW calculations⁴ and demonstrate that stishovite transforms to the CaCl₂ structure on quasi-hydrostatic compression—not only is the transition pressure closely predicted, but the calculated behaviour of the modes is accurate (Fig. 3). At pressures above the phase transition, four sharp bands and one broad band are observed in the Raman spectrum (Table 1, Fig. 3). Mode assignments for the sharp bands of the high-pressure polymorph were easily made by analogy with similar CaCl₂-structured compounds²¹. The origin of the broad, low-frequency band is discussed in the legends of Figs

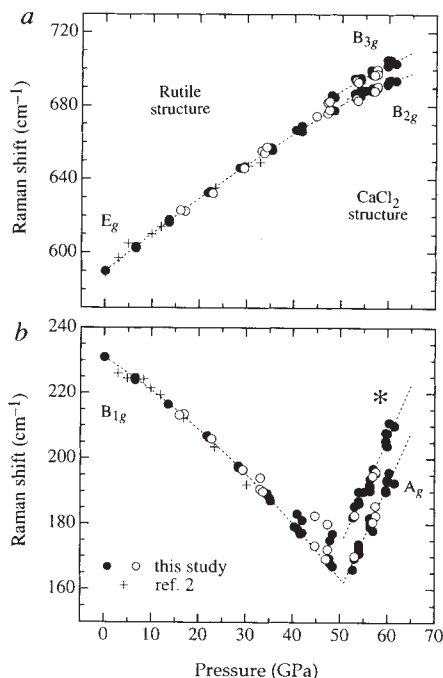


FIG. 2 Frequency shifts of the mid-frequency (a) and low-frequency (b) Raman bands during compression (filled circles) and decompression (empty circles) through the rutile-to-CaCl₂ transition in silica. The dashed curves are quadratic fits of the data plotted to guide the eye. The frequency shift of the new broad band (marked by asterisk) is found to parallel the frequency shift of the CaCl₂ A_g mode. Results at lower pressure are in excellent agreement with those reported in ref. 2.

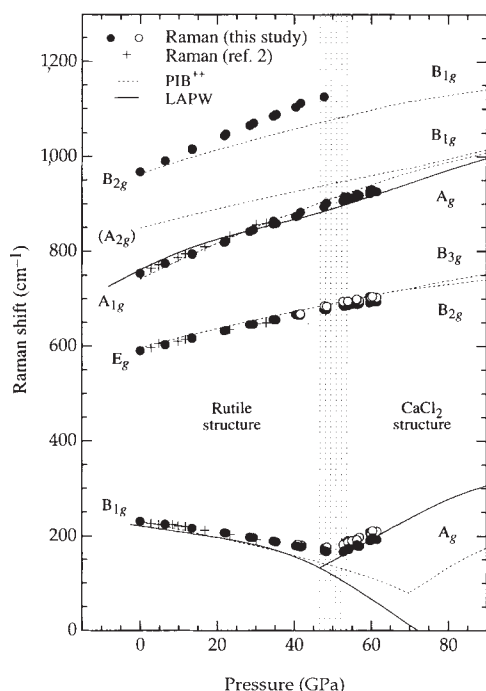


FIG. 3 Comparison of experimental frequency shifts during compression with theoretical LAPW⁴ and PIB⁺⁺ calculations (see text). The experimentally determined pressure for the rutile-to-CaCl₂ transition in silica from this work is shown as the stippled vertical band. LAPW results constraining the symmetry to remain that of stishovite appear as extension of the rutile B_{1g} frequency shift curve at pressures above the transition. The general mode behaviour is reproduced well by the new PIB⁺⁺ model; the frequency shifts of the rutile B_{2g} mode and especially the rutile E_g, which were not calculated in the LAPW study, are in good agreement with the Raman measurements, especially considering that there is no fit to experimental data. The PIB⁺⁺ model predicts that the Raman-active CaCl₂ B_{1g} mode that correlates with the silent (non-active) A_{2g} mode in stishovite is almost degenerate with the high-frequency CaCl₂ A_g mode. Thus it is likely that this CaCl₂ B_{1g} mode was hidden in the Raman measurements by the strong CaCl₂ A_g mode. Additionally, the PIB⁺⁺ results demonstrate that the new, broad feature observed at low frequency (Figs 1 and 2) cannot be either of the unobserved CaCl₂ B_{1g} modes, which should occur above 900 cm⁻¹. The transition pressure, which is very sensitive to the details of the potential, is over-estimated by 20 GPa by the PIB⁺⁺ calculations. Although this model uses the LDA and is non-empirical, the potential is parametrized and fitted to the LAPW total energies, and thus is less reliable than the electronic-structure calculations themselves. The observed flattening of the pressure dependence of the rutile B_{1g} mode and CaCl₂ low-frequency A_g mode relative to the zero-temperature LAPW calculations near the transition is probably due to anharmonicity, which was not included in the first-principles calculation.

1 and 3. The excellent agreement between theory and experiment demonstrates the power of new electronic-structure methods; this may be one of the best examples of an accurate first-principles prediction of a phase transition.

A pressure of 50 GPa corresponds to a depth in the Earth of ~1,200 km. The temperature^{22,23} of the mantle at this depth ranges from 2,000 to 2,500 K. We examined the temperature dependence of the transition in silica and also confirmed mode assignments using a new non-empirical lattice dynamical model (PIB⁺⁺)¹⁸, which is based on the potential induced breathing (PIB) model but is fitted to the results of the earlier LAPW frozen-phonon calculations⁴. Results at ambient temperature are shown in Fig. 3. Most importantly, the PIB⁺⁺ calculations predict a small temperature dependence for the transition of ~250 K GPa⁻¹, so that any free silica in the lower mantle would transform to the CaCl₂ structure at ~60 GPa assuming an upper temperature bound of 2,500 K.

Thermoelastic data indicate a high perovskite abundance in the lower mantle, and possible enrichment of silica and/or iron relative to the upper mantle^{22,23}. The existence of a stable silica phase with free energy lower than that of stishovite drives the reaction (Mg, Fe)SiO₃ perovskite → (Mg, Fe)O magnesio-wüstite + SiO₂. Additionally, if iron enrichment is significant,

the production of magnesio-wüstite and a silica phase is further promoted by the inability of perovskite to accommodate large amounts of iron²³.

We have shown that the rutile-to-CaCl₂ transition occurs in silica at lower-mantle pressures. Because an elastic instability is associated with the transition, a shear modulus is anomalously soft near the phase change. Free silica in the lower mantle may therefore lead to an observable seismic discontinuity around a depth of 1,200–1,500 km. Although globally recognized discontinuities have been restricted to the upper mantle (<660 km), there are reports of local discontinuities which cluster in the range of the new transition pressure^{24–26}. Thus although present seismological observations seem to preclude global discontinuities in the lower mantle²⁷, new techniques allowing for mapping at much higher resolutions should be employed to look for the seismic signal of the stishovite transformation. Such a signal, or lack thereof, would place limits on the amount of free silica in the lower mantle. Moreover, the new transition is geophysically important for experimentally documented reactions involving the production of free silica from iron and perovskite which may be associated with the anomalous seismic structure in D'', the 100- to 300-km region at the base of the mantle²⁸. □

Received 2 November 1994; accepted 17 January 1995.

- Nagel, L. & O'Keeffe, M. *Mater. Res. Bull.* **6**, 1317–1320 (1971).
- Hemley, R. J. in *High-Pressure Research in Mineral Physics* (eds Manghni, M. H. & Syono, Y.) 347–359 (Terra Scientific, Tokyo and Am. Geophys. Union, Washington DC, 1987).
- Tsuchida, Y. & Yagi, T. *Nature* **340**, 217–220 (1989).
- Cohen, R. E. in *High-Pressure Research: Application to Earth and Planetary Sciences* (eds Syono, Y. & Manghni, M. H.) 425–431 (Terra Scientific, Tokyo and Am. Geophys. Union, Washington DC, 1992).
- Weidner, D. J., Bass, J. D., Ringwood, A. E. & Sinclair, W. J. *geophys. Res.* **87**, 4740–4746 (1982).
- Liu, L. G., Bassett, W. A. & Takahashi, T. J. *geophys. Res.* **79**, 1160–1164 (1974).
- Sato, Y. *Earth planet. Sci. Lett.* **34**, 307–312 (1977).
- Sugiyama, M., Endo, S. & Koto, K. *Miner. J.* **13**, 455–466 (1987).
- Ross, N. L., Shu, J.-F. & Hazen, R. M. *Am. Miner.* **75**, 739–747 (1990).
- McMillan, P. F. & Hofmeister, A. M. in *Spectroscopic Methods in Mineralogy and Geology* (ed. Hawthorne, F. C.) 99–159 (Mineralogical Soc. Am., Washington DC, 1988).
- Lyons, R. J. P. *Nature* **196**, 266–267 (1962).
- Kieffer, S. W. *Rev. Geophys. Space Phys.* **17**, 20–33 (1979).
- Hofmeister, A. M., Xu, J. & Akimoto, S. *Am. Miner.* **75**, 951–955 (1990).
- Hemley, R. J., Mao, H. K. & Chao, E. C. T. *Phys. Chem. Miner.* **13**, 285–295 (1986).

- Williams, Q., Hemley, R. J., Kruger, M. B. & Jeanloz, R. J. *geophys. Res.* **98**, 22157–22170 (1993).
- Stishov, S. M. & Popova, S. V. *Chemistry (USSR)* **10**, 923–926 (1961).
- Chao, E. C. T., Fahey, J. J., Littler, J. & Milton, D. J. J. *geophys. Res.* **67**, 419–421 (1962).
- Cohen, R. E. in *Silica: Physical Behavior, Geochemistry and Materials Applications* (eds Heaney, P. J., Prewitt, C. T. & Gibbs, G. V.) 369–402 (Reviews in Mineralogy, Vol. 29, Mineralogical Soc. Am., Washington DC, 1994).
- Mao, H. K. & Bell, P. M. *Yb. Carnegie Instn Wash.* **77**, 904–908 (1978).
- Bell, P. B. & Mao, H. K. *Yb. Carnegie Instn Wash.* **80**, 404–406 (1981).
- Weber, W. H., Graham, G. W. & McBride, J. R. *Phys. Rev.* **B42**, 10969–10975 (1990).
- Knittle, E., Jeanloz, R. & Smith, G. L. *Nature* **319**, 214–216 (1986).
- Stixrude, L., Hemley, R. J., Fei, Y. & Mao, H. K. *Science* **257**, 1099–1101 (1992).
- Johnson, L. R. *Bull. seism. Soc. Am.* **59**, 973–1008 (1969).
- Vinnik, L. P., Lukk, A. A. & Nikolaev, A. V. *Phys. Earth planet. Inter.* **5**, 328–331 (1972).
- Wicks, C. W. *EOS (abstr.)* **74**, 550 (1993).
- Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **22**, 277–288 (1981).
- Knittle, E. & Jeanloz, R. *Science* **251**, 1438–1443 (1991).
- Jayaraman, A., Wood, D. L. & Maines, R. G. *Sr Phys. Rev.* **B35**, 8316–8321 (1987).
- Mao, H. K., Xu, J. & Bell, P. M. J. *geophys. Res.* **91**, 4673–4676 (1986).

ACKNOWLEDGEMENTS. We thank Y. Fei and M. Walter for synthesis of starting samples. This research was supported by the US NSF, including support to D. Veblen at the Department of Earth and Planetary Sciences of The Johns Hopkins University.

Confounding influence of magnetic fabric on sedimentary records of a field reversal

Xavier Quidelleur*, John Holt†
& Jean-Pierre Valet*

* Laboratoire de Paléomagnétisme,
Institut de Physique du Globe de Paris,
4, place Jussieu, 75252 Paris Cedex 05, France

† Division of Geological and Planetary Sciences,
California Institute of Technology, Pasadena, California 91125, USA

RECENT compilations of geomagnetic reversal records^{1–8} have generated a controversy as to whether the geomagnetic field is geographically biased during polarity transitions. At present there is general agreement that the virtual geomagnetic poles recorded from Cenozoic sediments preferentially lie over the Americas (or the antipodal longitude), yet such a preference is not statistically established^{7,9}. However, it is intriguing that the claimed preferred paths lie 90° away from the site longitude^{4,9}. Although this may be partly inherent in the very poor geographical distribution of the sites, we prefer not to rely on fortuitous coincidences. Several authors have argued that sedimentary palaeomagnetic records may be modified by artefacts linked to the acquisition of magnetization^{10–13}. Here we report that in two sedimentary records of the Upper Olduvai reversal from Confidence Hills, California, the declinations of the remanent magnetization recorded during the reversal are similar to the directions of the maximum horizontal axes of the ellipsoids of magnetic anisotropy. This supports the idea that, at times of low geomagnetic intensity (such as during a reversal), factors other than the geomagnetic field influence the orientation of elongated grains.

High-resolution records of the Upper Olduvai geomagnetic reversal were recently obtained from two adjacent sections of lacustrine deposits (Death March Canyon, hereafter referred to as section 1, and Confusion Canyon, or section 2) in the Confidence Hills of southern Death Valley, California¹⁴. These sediments are characterized by a high deposition rate, 33 and 28 cm kyr⁻¹ for sections 1 and 2, respectively. The difference in bedding attitude between the two sections (strike/dip for section 1 is N50° W/45° NE, whereas section 2 is N60° W/98° NE) allowed the first fold test¹⁴ of transitional directions and was used to recognize present-day field overprints. It was thus possible to assess the primary origin of the characteristic component of magnetization carried by single-domain titanomagnetite. The virtual geomagnetic pole (VGP) paths, which cluster over the Atlantic and western Pacific oceans, show no consistency with other records of the same transition. The most obvious characteristic is the tendency for VGPs to lie ~90° away from the site meridian. It has been suggested^{9,15} that this feature, commonly observed in sedimentary records, could reflect the tendency for elongated particles to be deposited with their long axes aligned parallel to the horizontal plane when the field strength is not large enough, a process that would result in low magnetic inclinations which in turn cause deviations of the VGP paths from the site meridian. We thus found it useful and interesting to investigate the relationship between the magnetic fabric and the directions of the remanent magnetization for transitional and non-transitional samples from the Confidence Hills. Because of the high level of detail present in these records, they are particularly suitable for studies of acquisition processes of magnetization in sediments during transitions.

The anisotropy of magnetic susceptibility (AMS)^{16–21} has long been used as a method for quantifying sedimentary fabrics. However, this technique is most sensitive to coarse magnetite

grains which do not carry the characteristic remanent magnetization. It is therefore difficult to interpret AMS data when dealing with the anisotropy of the carriers of magnetic remanence. This problem can be overcome by measuring the anhysteretic remanent magnetization (ARM) which is carried preferentially by single-domain and pseudo-single-domain grains^{22–24}. The ARM anisotropy (AAR), using the ARM acquisition ability of a given sample²⁴, has been shown to be a useful complement to standard AMS measurements. In addition, there is no contribution to AAR from paramagnetic and diamagnetic grains. We therefore used AAR in addition to AMS in order to fully characterize the magnetic fabric.

Oriented cylindrical samples (74 in total: 41 from section 1 and 33 from section 2) were collected across the Upper Olduvai transition zone. The measurements of AMS were performed at the laboratory of Parc St Maur (France) with a Kappabridge KLY-2 (Geofyzika Brno). The AAR was determined from measurements performed in the shielded room of the Institut de Physique du Globe de Paris using the following procedure: (1) the specimen was demagnetized in a peak alternating field of 90 mT; (2) an ARM was imparted along a specific sample axis in a peak alternating field of 80 mT with a direct field of 0.1 mT; (3) the ARM was partially demagnetized with a peak alternating field of 20 mT along the same sample axis; (4) the ARM was measured with a CTF cryogenic magnetometer using a multi-orientation sample holder so that the induced ARM could always be measured along the *z*-axis of the magnetometer; (5) the sample was then demagnetized along three axes with a peak alternating field of 90 mT in order to remove the previous ARM; (6) another ARM was then imparted along a different axis of the sample. This procedure was repeated for nine different orientations using the same conventions as for the AMS determinations. The best-fit anisotropy tensor and its eigenvalues and eigenvectors were calculated using the method of Jelinek²⁵. This allows a precise determination of the anisotropy tensor. To rule out any secondary effect, the very small component remaining after demagnetization was subtracted from any imparted ARM.

The maximum ($K_{\max}$) and minimum ($K_{\min}$) axes of AMS and AAR are shown in equal-area stereograms in Fig. 1. Both sections show a well defined AMS foliation with the minimum axes perpendicular to the bedding plane and maximum AMS axes with a tendency for northwest–southeast preferential orientation. The maximum AAR axes are also parallel to the bedding plane. However, the AAR $K_{\max}$ declinations from section 1 are quite scattered whereas they show a northwest–southeast lineation close to the bedding strike in section 2. The $K_{\min}$ axes are vertical in section 1 but deviate from 90° in section 2. The percentages of anisotropy of the AAR are greater than for those of AMS ($H = 9.9 \pm 5.4$ and 7.9 ± 4.3 for sections 1 and 2 respectively versus 3.8 ± 1.9 and 2.5 ± 1.4 , where $H = 100((K_{\max} - K_{\min})/K_{\text{int}})$ and K_{int} is the intermediate axis. The values of the shape parameter for the AAR ($F = (K_{\max} - K_{\text{int}})/(K_{\max} - K_{\min})$) are typical of oblate ellipsoids ($F = 0.31 \pm 0.22$) in section 1 and closer to prolate structures ($F = 0.47 \pm 0.19$) in section 2. These anisotropy parameters show no apparent correlation with directional changes.

Thus there are obviously some differences between the two sections. In the absence of deformation, the AMS in sediments²⁶ is linked to the shape of the elongated grains that come to rest during deposition with their long axes parallel to the bedding plane and eventually azimuthally oriented by bottom currents. Shape anisotropy is common for single-domain or pseudo-single-domain grains of magnetite and is reflected by the AAR²⁷. The fact that the inclinations of the $K_{\min}$ axes (of both anisotropies) remain tightly grouped close to 90° in section 1 whereas the $K_{\max}$ are widely scattered (Fig. 1) in the bedding plane indicates that the original depositional fabric has been preserved, unaffected by deformation. Although this seems to be much less apparent in the AMS, the dispersion of the AAR $K_{\min}$ axes in

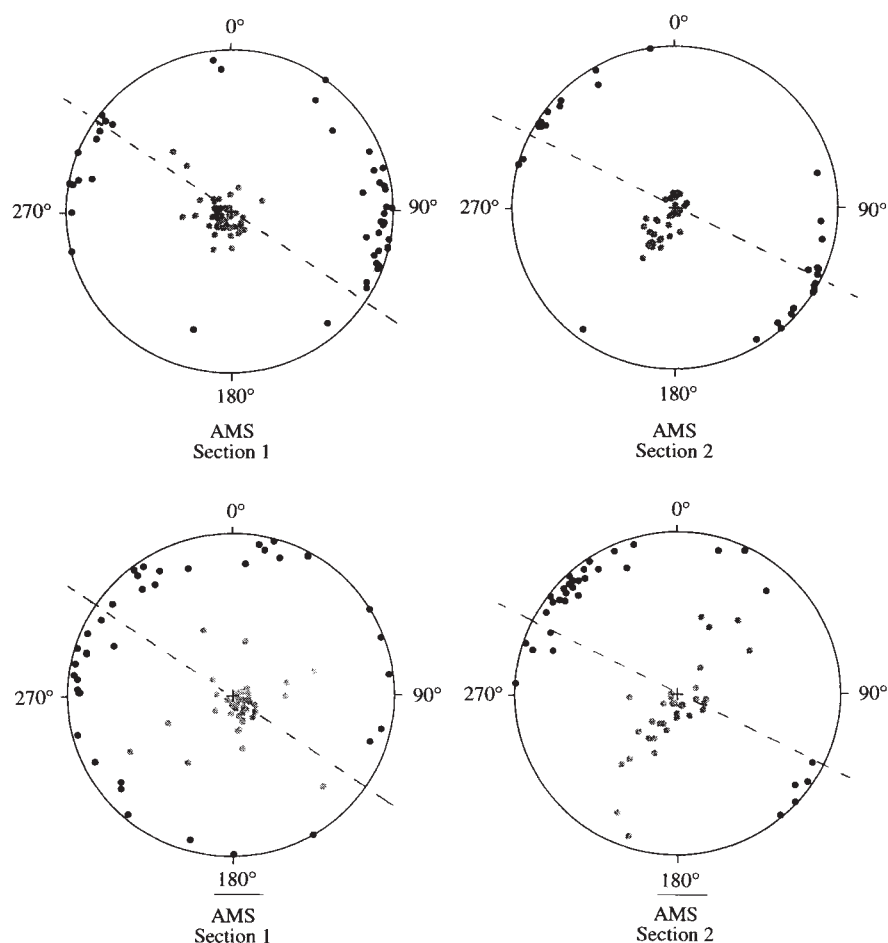


FIG. 1 Top, equal-area stereo plots of the directions of the $K_{\max}$ (black dot) and $K_{\min}$ axes (grey dot) of magnetic susceptibility (AMS) for samples from section 1 (left) and section 2 (right). Bottom, the $K_{\max}$ and $K_{\min}$ axes of the anhyseretic remanence (AAR). Dotted lines show the directions of the bedding strike.

section 2 and the tendency of the $K_{\max}$ axes to lie closer to the direction of the bedding strike (300°) indicates that deformation could be more significant than in the previous section. Indeed, in presence of intense folding the orientation of the $K_{\max}$ axes is expected to be parallel to the fold axis while the inclinations tend to be scattered away from the horizontal plane^{28,29}. The anisotropy measurements (especially the AAR) are thus quite consistent with the field observations which show that section 1 (45° dip) experienced less folding than section 2 (98°). On the other hand, these differences in AAR $K_{\max}$ scatter could be explained by the different sedimentary environments characterizing these sections¹⁴. Section 2 is a high-energy environment whereas section 1 has quieter depositional conditions which would produce more scattered directions of the AAR $K_{\max}$ axes. One can speculate that both of these mechanisms of non-magnetic origin should have a limited effect on the remanence directions outside the transition. This is because they mostly influence orientation of elongated grains whereas remanence is carried primarily by equant grains in these sediments¹⁴.

In Fig. 2 we compare the directions of the AAR $K_{\max}$ axes with the declinations of the remanence. $K_{\max}$ has no polarity, and hence the axes define two declinations 180° apart. We plotted the declinations of $K_{\max}$ with the minimum deviation from the direction of the remanence. The most striking feature is that in both sections the declinations of the AAR $K_{\max}$ axes more closely match the remanence declinations during the transition (220–228 m in section 1 and 167–169 m in section 2) than outside this interval. Furthermore, it can be seen that there is a good agreement if only directions associated with a low-latitude VGP are considered. There are several possible interpretations of the apparent correlation between the transitional declinations and the orientation of the AAR $K_{\max}$ axes. Significant effects of

smoothing of the transitional directions by post-depositional reorientations of the grains seem unlikely because of the very high deposition rates and the large directional jumps observed during both transitional intervals. The fold test also eliminates the possibility of overprinting by the present geomagnetic field. The AAR could reflect preferential orientation of the long axes of single-domain and/or pseudo-single-domain grains by the geomagnetic field. But in this case, the declinations and the inclinations of the AAR axes should also more or less reflect those of the remanence, particularly outside the transition interval where the geomagnetic field intensity is much higher. In both sections the inclinations of the $K_{\max}$ axes remain close to 0° and thus strikingly different from those of the remanence outside the transitional intervals. It therefore seems difficult to invoke only a geomagnetic origin for the AAR.

We may first find it surprising that under such conditions similar directional changes have been recorded at both sections. Although there is an overall similarity between the two records, there are large swings and abrupt variations of declination in section 2 which contrast with the much smoother behaviour of the declination in section 1 (Fig. 2). In both cases, this different character is duplicated by the declinations of the AAR $K_{\max}$ axes. The similarity between the orientations of the transitional directions and the $K_{\max}$ axes of the AAR suggests that the anisotropy effects may have played a significant role in the records of the transitional directions.

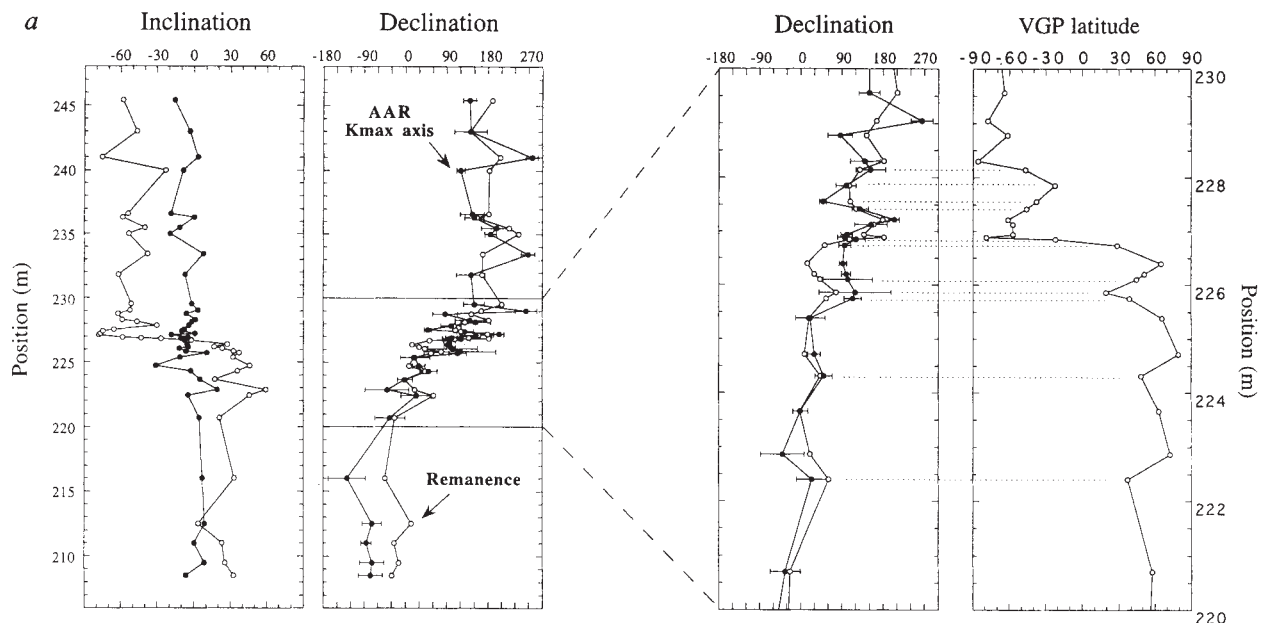
One may thus wonder in this case why the orientations of the $K_{\max}$ axes of the AAR do not match those of the characteristic component of magnetization outside the transition zone. Also puzzling is the fact that the inclinations of the remanence do not match exactly those of the $K_{\max}$ axes during the transition, although it is noticeable that the largest similarities occur for

transitional directions with VGPs at the lowest latitudes. But if we assume that the orientations of the elongated grains dominate the remanent magnetization during the transition, this should not necessarily be so outside the reversal because most of the characteristic remanent magnetization is carried by single domain equant grains without shape anisotropy. This is supported by transmission electron microscope observations of fine-grained separates that confirm the presence of equant single-domain grains of titanomagnetite and detect also the presence of non-equant grains¹⁴. During periods of stable polarity, the fraction of elongated grains lying with their long axes parallel to the bedding eventually introduces a bias into the remanence which is usually described as an inclination error³⁰ and is observed at Confidence Hills¹⁴. When the external field becomes

weak, as during a geomagnetic reversal, the alignment of the equant particles by the geomagnetic field becomes much weaker so that the contribution of the fraction of elongated particles to the remanence may be considerably enhanced if not dominant. The resultant magnetization is thus much more sensitive to grains with their moments more or less parallel to the horizontal plane. This process induces significant inclination shallowing, which in turn causes significant deviations of the VGP paths (between 60° and 120°) from the site meridian⁹. It can be noted that this hypothesis is slightly different from one previously proposed⁹, in which only elongated grains were involved.

Although we cannot preclude the possibility that these results could be peculiar to these sediments, they imply some degree of inability of sediments in general to accurately record the field

Section 1



Section 2

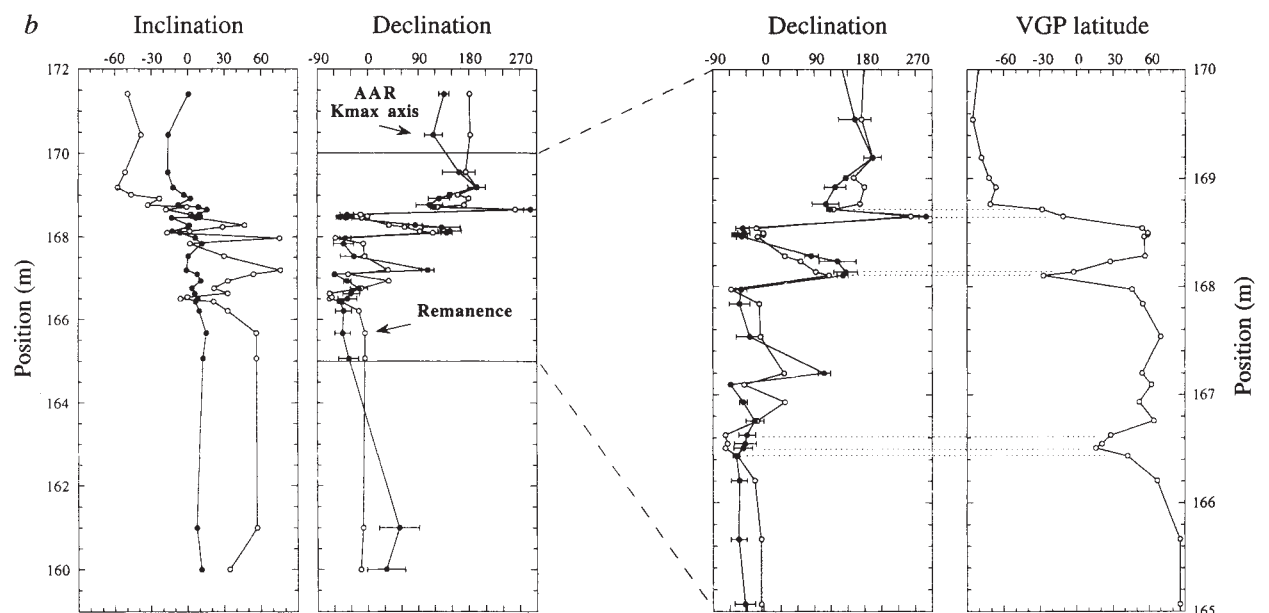


FIG. 2 a, The two left-hand plots show inclinations and declinations (in degrees) of the AAR $K_{\max}$ axes (filled circles) compared with the declinations of the remanence (open circles) as a function of the stratigraphic

height in section 1 deposits. The two right-hand plots show an enlarged view of the declinations and VGP latitudes across the transition zone. b, as a but for section 2 deposits.

direction during periods of low field intensity. This clearly emphasizes the importance of anisotropy measurements before any interpretation of transitional directions in terms of geomagnetic field behaviour can be carried out. □

Received 21 September 1994; accepted 3 February 1995.

1. Clement, B. M. *Earth planet. Sci. Lett.* **104**, 48–58 (1991).
2. Tric, E. et al. *Phys. Earth planet. Inter.* **65**, 319–336 (1991).
3. Laj, C., Mazaud, A., Fuller, M., Weeks, R. & Herrero-Bervera, E. *Nature* **351**, 447 (1991).
4. Valet, J. P., Tucholka, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400–407 (1992).
5. Hoffman, K. A. *Nature* **359**, 789–794 (1992).
6. Bogue, S. W. & Merrill, R. T. *Rev. Earth planet. Sci.* **20**, 181–219 (1992).
7. McFadden, P. L., Barton, C. E. & Merrill, R. T. *Nature* **361**, 342–344 (1993).
8. Prévot, M. & Camps, P. *Nature* **366**, 53–57 (1993).
9. Quidelleur, X. & Valet, J. P. *Phys. Earth planet. Inter.* **82**, 27–48 (1994).
10. Rochette, P. *Earth planet. Sci. Lett.* **98**, 33–39 (1990).
11. Van Hoof, A. A. M. & Langereis, C. G. *Nature* **351**, 223–225 (1991).
12. Langereis, C. G., van Hoof, A. A. M. & Rochette, P. *Nature* **358**, 226–230 (1992).
13. Kent, D. V. & Schneider, D. A. C. *EOS (abstr.)* **75**, 119 (1994).
14. Holt, J. W. & Kirschvink, J. L. *Earth planet. Sci. Lett.* (submitted).

15. Chauvin, A., Roperch, P. & Duncan, R. A. *J. geophys. Res.* **95**, 2727–2752 (1990).
16. Ising, G. *Astron. Uch Fysik.* **29A**, 1–37 (1942).
17. Graham, J. W. *Geol. Soc. Am. Abstr. Progr.* **65**, 2257 (1954).
18. Granar, L. *Ark. Geofys.* **3**, 1–40 (1958).
19. Rees, A. I. *Sedimentology* **4**, 257–271 (1965).
20. Hamilton, N. & Rees, A. I. *Paleogeophysics* (ed. Runcorn, S. K.) 445–464 (Academic, New York, 1970).
21. Kent, D. V. & Lowrie, W. *Earth planet. Sci. Lett.* **28**, 1–12 (1975).
22. Banerjee, S. K., King, J. & Marvin, J. *Geophys. Res. Lett.* **8**, 333–336 (1981).
23. King, J., Banerjee, S. K., Marvin, J. & Özdemir, Ö. *Earth planet. Sci. Lett.* **59**, 404–419 (1982).
24. McCabe, C., Jackson, M. & Ellwood, B. B. *Geophys. Res. Lett.* **12**, 333–336 (1985).
25. Jelinek, V. *The Statistical Theory of Measuring Anisotropy of Magnetic Susceptibility of Rocks and its Application* (Geofyzika, Brno, 1977).
26. Ellwood, B. B. *Mar. Geol.* **34**, 83–90 (1980).
27. Jackson, M., Gruber, W., Marvin, J. & Banerjee, S. K. *Geophys. Res. Lett.* **15**, 440–443 (1988).
28. Hrouda, F. *Geophys. Surv.* **5**, 37–82 (1982).
29. Borradaile, G. *Tectonophysics* **156**, 1–20 (1988).
30. Opdyke, N. D. & Henry, K. W. *Earth planet. Sci. Lett.* **6**, 138–151 (1969).

ACKNOWLEDGEMENTS. We thank J. P. Cogné, C. Laj and C. Richter for helpful suggestions and discussions, B. Henry for help with measurements and interpretations of the AMS, and R. Warner and C. Scrivner for assistance with field work. This work was supported by the French program CNRS INSU DBT Terre Profonde and the US NSF.

Geodetic observations of very rapid convergence and back-arc extension at the Tonga arc

Michael Bevis*, F. W. Taylor†, B. E. Schutz‡, Jacques Recy§, B. L. Isacks||, Saimone Helu¶, Rajendra Singh*, Eric Kendrick**, James Stowell††, Brian Taylor* & Stephane Calmant‡‡

* Hawaii Institute of Geophysics and Planetology, 2525 Correa Rd, University of Hawaii, Honolulu, Hawaii 96822, USA

† Institute for Geophysics, University of Texas at Austin, Austin, Texas 78759, USA

‡ Center for Space Research, University of Texas at Austin, Austin, Texas 78712, USA

§ Institut Français de Recherche Scientifique Pour le Développement en Coopération (ORSTOM), Villefranche sur Mer, France

|| Department of Geological Sciences, Cornell University, Ithaca, New York 14853, USA

¶ Ministry of Lands, Survey and Natural Resources, Nuku'alofa, Tonga

§§ Fiji Mineral Resources Department, Suva, Fiji

** Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina 27695, USA

†† Universities Navstar Consortium, PO Box 3000, Boulder, Colorado 80309, USA

‡‡ Institut Français de Recherche Scientifique Pour le Développement en Coopération (ORSTOM), Noumea, New Caledonia

THE Earth's most active zone of mantle seismicity arises from the subduction of the Pacific plate at the Tonga trench¹. It is not known why this slab generates so many more earthquakes than other subducting slabs worldwide. Above the subduction zone the active Tofua (Tonga) volcanic arc is separated by the V-shaped Lau basin from a remnant arc, the Lau ridge, located at the eastern edge of the Australian plate². The irregular and discontinuous magnetic lineations within the basin have proven difficult to interpret^{3,4}, and so the regional kinematic framework has been obscure. We report geodetic measurements of crustal motion within the Tonga–Lau system, which reveal the fastest crustal motions yet observed. The Lau basin is opening at a rate which increases northwards to a maximum of $\sim 160 \text{ mm yr}^{-1}$. No straining is observed within the northern Tonga ridge, suggesting that it comprises part of a rigid microplate. Convergence rates across the Tonga trench increase northwards to a maximum of $\sim 240 \text{ mm yr}^{-1}$. The extraordinary seismic activity of the subducting slab is probably related to this unusually rapid subduction.

It is widely believed that the Lau basin opened as the Tonga ridge rifted and drifted away from the Lau ridge (Fig. 1), but no consensus exists as to how the crust of the Lau basin was formed. For more than twenty years, geologists and geophysicists have debated the relative importance of localized sea-floor spreading (and microplate tectonics) versus distributed deformation within the Lau and North Fiji basins^{4,5}. Recent drilling and marine geophysical surveys within the Tonga–Lau system have prompted a hybrid model in which the Lau basin began opening $\sim 6 \text{ Myr}$ ago by diffuse Basin-and-Range style rifting, after which organized sea-floor spreading initiated in the north $\geq 4 \text{ Myr}$ ago and continues to propagate southwards³. But Hamburger and Isacks⁵ have argued that the contemporary seismicity of the Lau and North Fiji basins is quite unlike that of spreading centres at mid-ocean ridge systems, and instead implies that these interarc regions are now opening by a diffuse and shear-dominated system of deformation. It is possible that diffuse deformation and localized sea-floor spreading may now coexist within the basin³.

The Global Positioning System⁶ (GPS) was used to measure the relative geometry of a network of monuments during eight-day observation sessions that took place in July 1990 and July 1992. During each session we occupied nine stations (Fig. 1, Table 1) with dual-frequency geodetic GPS receivers for 6–8 consecutive days, except station LLAU which we occupied for 3 days in 1990 and 4 days in 1992. The GPS data were analysed using TEXGAP software⁷, and the daily geodetic solutions were combined to estimate instantaneous network geometries for each field campaign. These epoch geometries were compared to estimate station velocities (Table 1, Fig. 1) in a reference frame which minimizes the motions of the three sites in the Pacific plate. Because the GPS measurement accuracies improved between 1990 and 1992, about three-quarters of the uncertainty in the velocity solutions derives from measurement errors in the 1990 campaign.

The GPS station velocities in this Pacific-fixed frame (Fig. 1, Table 1) fall into three groups. (1) The Pacific-plate sites (RARO, NIUE, WSAM) have no significant motion. There is no significant relative motion of these sites in any reference frame. (2) The Fiji–Lau sites (VITI, VANU, LLAU) are moving at $\sim 80 \text{ mm yr}^{-1}$ (bearing $\sim \text{N}91^\circ\text{E}$), which is very similar to the Australia–Pacific convergence rates predicted by the global plate motion model NUVEL-1A⁸ (Table 2), implying little or no motion of Fiji relative to the Australian plate. We cannot resolve whether these stations lie within the rigid core of the Australian plate, or within a marginal zone of (Australian) intraplate deformation associated with the North Fiji fracture zone. (3) The velocities of the three Tongan sites (TGPU, VAVA, NTPT) increase northwards from 164 to 240 mm yr^{-1} , with azimuths $\sim \text{N}106^\circ\text{E}$ which do not differ significantly from the mean slip

TABLE 1 GPS site velocities in a Pacific-fixed reference frame

"Fixed" station	Station code	Horizontal velocity (mm yr ⁻¹)	"Moving" station	Station code	Horizontal velocity (mm yr ⁻¹)	Velocity azimuth (N°E)
Rarotonga	RARO	5 ± 12	Niutoputapu	NTPT	240 ± 11	104 ± 1
Upolu (W. Samoa)	WSAM	1 ± 8	Vava'u	VAVA	205 ± 10	108 ± 1
Niue	NIUE	5 ± 16	Tongatapu	TGPU	164 ± 5	105 ± 2
			Vanua Levu	VANU	79 ± 15	92 ± 2
			Viti Levu	VITI	86 ± 14	94 ± 2
			Lakeba (Lau)	LLAU	78 ± 11	89 ± 2

Velocities of nine GPS stations (part of the southwest Pacific GPS network) in a reference frame which minimizes the horizontal motions of the three stations within the Pacific plate (1990–92). Uncertainties stated are standard errors. Note that there is no significant motion of the three stations (RARO, WSAM, NIUE) that lie within the Pacific plate, indicating that this reference frame can be considered Pacific-fixed.

vector azimuths associated with shallow underthrusting earthquakes near the Tongan stations⁹. The GPS site velocities in Tonga (Table 1) differ substantially in magnitude and azimuth from the Australia–Pacific convergence vectors (Table 2). No large earthquakes occurred near the Tonga trench between 1990 and 1992. Unless interseismic convergence across a subduction zone can be very unsteady, we must conclude that the velocities of the GPS sites in Tonga represent the geological subduction rate. The rates of convergence across the trench at the latitudes of Vava'u and Niutoputapu represent the highest secular plate velocities ever observed. Given the very low seismic coupling of this region¹⁰ (that is, very low ratios of seismic to tectonic slip) it is likely that most if not all of the convergence observed using GPS was achieved by aseismic creep.

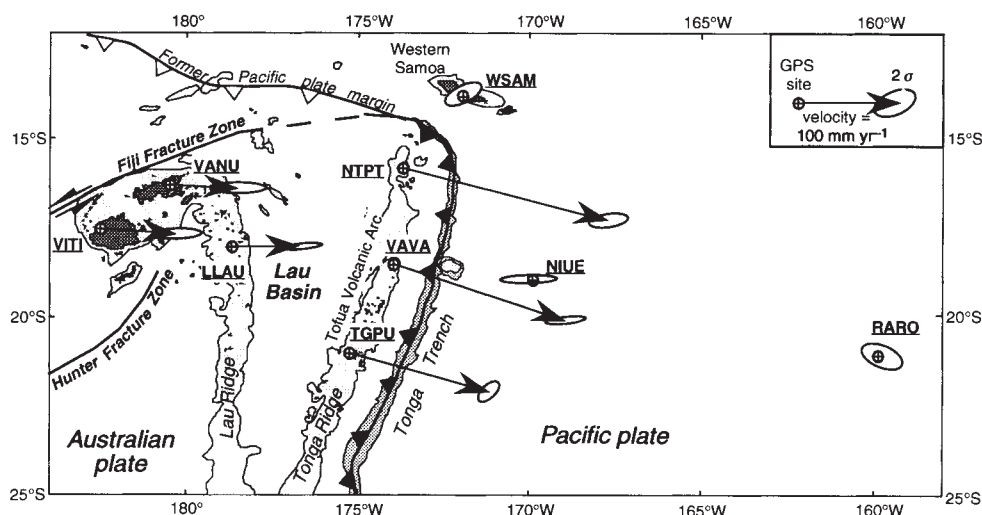
The subduction rates in Tonga result from a combination of Australian–Pacific plate convergence and the opening of the Lau basin (Australia–Tonga divergence). We can isolate this second component by using NUVEL-1A to switch to an Australia-fixed frame of reference. (At each station, one subtracts the motion of Australia relative to the Pacific according to NUVEL-1A from the GPS velocities already described). In this Australia-fixed frame (Fig. 2) the stations in Fiji and northern Lau have little or no significant motion, and the velocities of the Tongan stations (Table 3) characterize the motion of the Tonga ridge relative to the Lau ridge due to extension within the intervening

Lau basin. These velocities increase northwards from 91 to 159 mm yr⁻¹. It has been proposed that the central Lau basin has converted from diffuse rifting to sea-floor spreading at the Eastern Lau and Central Lau spreading centres (Fig. 2), but the suggested rates of spreading^{3,11,12} in this area (55–99 mm yr⁻¹) fall well short of the GPS velocities (91–130 mm yr⁻¹). This suggests that either (1) spreading rates have substantially increased in the recent geological past and this change cannot be resolved in existing magnetic data, (2) the magnetic lineations have been misinterpreted, or (3) diffuse Basin-and-Range style deformation contributes significantly to the net basin extension rate, and localized sea-floor spreading supplements rather than replaces this older mode of extension. The last interpretation is supported by diffuse shallow seismicity within the Lau basin, including several events with focal mechanisms indicative of shear faulting⁵.

Despite the evidence for diffuse deformation within the Lau basin, our GPS results indicate no significant longitudinal extension of the arc between stations TGPU and NTPT. We suggest that the Tonga ridge (north of TGPU and probably north of 24° S) is behaving like a rigid windscreen wiper as it rotates to accommodate widening of the Lau basin.

The intense mantle seismicity beneath Tonga–Lau^{13,14} may arise because this area has the highest subduction and back-arc extension rates on Earth. Rapid subduction could enhance

FIG. 1 Crustal velocities in a Pacific-fixed reference frame derived from GPS measurements made in July 1990 and July 1992 at nine geodetic stations within the Tonga–Lau system. The velocity vectors are indicated by arrows; a 100 mm yr⁻¹ vector is shown in the inset (top-right corner). The Tonga and Lau ridges are represented using the 1,500-m isobath, and the Tonga trench using the 6,000-m isobath. The Pacific plate subducts westward at the Tonga trench and sinks into the mantle beneath the Tonga ridge and the V-shaped Lau basin. The active Tofua volcanic arc lies at the western edge of the Tonga ridge between about 18° and 22° S. The Lau ridge is a remnant arc that lies at the eastern edge of the Australian plate. North of this ridge the relative motion of the Pacific and Australian plates is accommodated mainly along the seismically active Fiji fracture zone. The Hunter fracture zone between about 174° and 178° E appears to be seismically inactive, suggesting that the Fiji platform, which includes stations VITI and VANU, is stationary or moves very slowly relative to the Australian plate. The GPS velocities (Table 1) are presented in a frame of reference which minimizes motions of the three stations (RARO, WSAM, NIUE) located within the Pacific plate. The



uncertainty in each velocity vector is represented by a 2σ (~86% confidence) error ellipse. The aspect ratio and orientation of these ellipses mainly reflects the fact that GPS satellite orbital errors are larger across-track than along-track. The velocities of the Tongan stations, which range from 164 ± 5 mm yr⁻¹ at Tongatapu (TGPU) to 240 ± 11 mm yr⁻¹ at Niutoputapu (NTPT), manifest the rate at which the Pacific plate is subducting.

TABLE 2 Australia-Pacific convergence rates according to NUVEL-1A

Station code	Horizontal velocity (mm yr ⁻¹)	Velocity azimuth (N°E)
NTPT	84 ± 1	93 ± 1
VAVA	79 ± 1	93 ± 1
TGPU	75 ± 1	92 ± 1
VANU	83 ± 1	88 ± 1
VITI	81 ± 1	87 ± 1
LLAU	80 ± 1	90 ± 1

Velocity of the Australian plate in a Pacific-fixed reference frame, predicted by the model NUVEL-1A, as evaluated at the GPS stations in Fiji and Tonga. Uncertainties stated are standard errors.

TABLE 3 Velocities of the Tongan GPS stations in an Australia-fixed reference frame

Station code	Horizontal velocity (mm yr ⁻¹)	Velocity azimuth (N°E)
NTPT	159 ± 10	110 ± 2
VAVA	130 ± 10	117 ± 3
TGPU	91 ± 05	115 ± 4

These velocities are obtained by subtracting Australian-Pacific plate convergence rates predicted by the global plate model NUVEL-1A from the Pacific-fixed GPS velocities (Table 1). Uncertainties stated are standard errors.

seismicity in several (possibly interrelated) ways. The Pacific plate entering the Tonga subduction zone is 100–140 Myr old, and therefore relatively cool, and as the plate is sinking into the mantle unusually rapidly it follows that it will be cooler, at a given depth, than other slabs worldwide, and therefore better able to store the elastic energy that is released in earthquakes¹⁵. The rate of seismic energy release depends on strain rate as well as seismogenic volume and temperature¹⁶. The migration and change in shape of the Tonga trench may account for the contorted nature of the subducting slab^{17,14}. Rapid changes in the curvature of a subducting slab imply high rates of bending¹⁸ (or unbending¹⁹) and membrane²⁰ strain. Rapid subduction towards a barrier (near 680 km depth) that resists or prevents slab penetration²¹ might also lead to higher-than-normal strain rates and seismic activity. If mantle earthquakes are caused by trans-

formational faulting²², rapid subduction might enhance mantle seismicity by way of the high volumetric flux of material undergoing transformation.

The rates of crustal motion revealed by GPS in the Tonga-Lau system are so high that it is natural to question the credibility of results derived from just two GPS field campaigns. We offer the following reassurances. (1) The formal errors associated with each GPS campaign solution are based on repeatability, but analyses of the 1990 and 1992 observations included external comparisons with very-long-baseline interferometry and satellite laser ranging measurements (made outside the southwest Pacific), and these comparisons indicate horizontal accuracies comparable to the formal error estimates⁷. (2) Some GPS site velocities can be tested against previous expectations. Most significantly, the GPS results indicate no significant relative motion of the three stations located within the Pacific plate, as one would expect. In no way is this null result built into the solution. Additionally, our solutions indicate little or no motion of the Fiji platform relative to Australia, as expected²³. (3) Four of the stations (RARO, WSAM, TGPU and VAVA) were observed in a small GPS campaign mounted in July 1988, and for this subset of stations it is possible to assess the consistency of the 1988 measurements with those made in 1990 and 1992. Although the experimental uncertainties in 1988 were much larger than those in 1990 and 1992, so that the 1988 measurements are now of marginal utility, they are statistically consistent with the trends established by the 1990 and 1992 observations. □

Received 21 September 1994; accepted 14 January 1995.

1. Oliver, J. O. & Isacks, B. L. *J. geophys. Res.* **72**, 4259–4275 (1967).
2. Karig, D. E. *J. geophys. Res.* **75**, 239–254 (1970).
3. Parsons, L. M. & Hawkins, J. W. *Proc. ODP Sci. Res.* **135**, 819–828 (1994).
4. Lawver, L. & Hawkins, J. W. *Tectonophysics* **45**, 323–339 (1978).
5. Hamburger, M. & Isacks, B. L. *Nature* **332**, 599–604 (1988).
6. Dixon, T. *Rev. Geophys.* **29**, 249–276 (1991).
7. Schütz, B. et al. *Bull. Géodésique* **67**, 224–240 (1993).
8. DeMets, C., Gordon, R.G., Argus, D. F. & Stein, S. *Geophys. Res. Lett.* **21**, 2191–2194 (1994).
9. Pelletier, B. & Louat, R. *Tectonophysics* **165**, 237–250 (1989).
10. Jarrard, R. D. *Rev. Geophys.* **24**, 217–284 (1986).
11. Weissel, J. K. in *Island Arcs, Deep Sea Trenches, and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C.) 429–436 (Am. Geophys. Union, Washington DC, 1977).
12. Morton, J. & Pohl, W. *Geol. Jb.* **92**, 93–108 (1990).
13. Frohlich, C. A. *Rev. Earth planet. Sci.* **17**, 227–254 (1989).
14. Hamburger, M. & Isacks, B. L. *J. geophys. Res.* **92**, 13841–13854 (1987).
15. Molnar, P., Freedman, D. & Shih, J. *Geophys. J. R. astr. Soc.* **56**, 41–54 (1979).
16. Bevis, M. *Science* **240**, 1317–1319 (1988).
17. Billington, S. thesis, Cornell Univ. (1980).
18. Chapple, W. M. & Forsyth, D. W. *J. geophys. Res.* **84**, 6729–6749 (1979).
19. Isacks, B. L. & Barazangi, M. in *Island Arcs, Deep Sea Trenches, and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C.) 99–114 (Am. Geophys. Union, Washington DC, 1977).
20. Bevis, M. *Nature* **323**, 52–53 (1986).
21. Giardini, D. *Phys. Earth planet. Inter.* **74**, 75–88 (1992).
22. Kirby, S. H. *J. geophys. Res.* **92**, 13789–13800 (1987).
23. Louat, R. & Pelletier, B. *Tectonophysics* **167**, 41–55 (1989).

ACKNOWLEDGEMENTS. We thank the Universities Navstar Consortium for equipment and logistical support, and the government agencies within Fiji, Tonga, Western Samoa, Niue and the Cook Islands for their assistance. We also thank M. Hamburger and D. Argus for comments and suggestions. This work was supported by the US NSF.

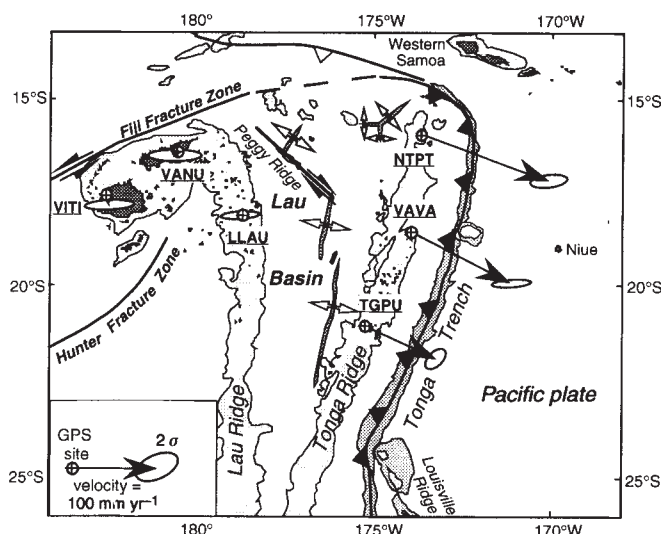


FIG. 2 Crustal velocities in an Australian-fixed reference frame realized by combining the GPS results and the Australian-Pacific plate motion predicted by model NUVEL-1A (see also Table 3). Also shown³ are the Eastern Lau spreading centre (west of station TGPU), the Central Lau spreading centre (between stations VAVA and LLAU), the Northwest Lau spreading centre (northeast of and adjacent to the Peggy ridge) and the Mangatolu triple junction (just west of station NTPT). The V-shaped Lau basin terminates just south of 24°S, near where the Louisville ridge intersects the Tonga trench. In this Australia-fixed frame the stations VITI, VANU and LLAU have little or no significant motion, while the velocities of the Tongan stations increase northwards from 91 ± 4 mm yr⁻¹ at TGPU to 159 ± 10 mm yr⁻¹ at NTPT. The Tonga ridge is rotating clockwise at ~7° Myr⁻¹ about a pole located close to the southern termination of the Lau basin, in order to accommodate widening of the basin achieved by localized sea-floor spreading and/or diffuse extension.

Interactive effects of ambient ozone and climate measured on growth of mature forest trees

S. B. McLaughlin & D. J. Downing

Environmental Sciences Division, Oak Ridge National Laboratory,
PO Box 2008, Oak Ridge, Tennessee 37831-6034, USA

GLOBAL increases in concentrations of tropospheric ozone are of worldwide concern because of its potential to affect both human and ecological health¹. Ozone has been implicated in the declining health of some forest tree species^{2,3} because of its effects on many growth-related processes in controlled studies³⁻⁵. Despite strong evidence that growth of forest tree seedlings can be reduced by ambient levels of ozone⁶, there has been little basis for evaluating the threshold and magnitude of direct effects on growth of mature, economically important forest trees. We describe here a five-year study of serial changes in stem circumference of 28 mature loblolly pine (*Pinus taeda* L.) trees. This study has defined a rough ozone response threshold and quantified short- and longer-term components of growth responses to varying ozone and climate variables. Ozone exposures at $\geq 40 \text{ nl l}^{-1}$ interacted with low soil moisture and high air temperatures to reduce short-term rates of stem expansion. Annual growth was also inversely related to seasonal ozone exposure and soil moisture stress. Effects varied widely between individual trees and years. Future ozone effects on forests are likely to be influenced by climate change and by projected increases in regional ozone pollution in industrialized countries.

Loblolly pine is an ecologically and economically important component of southern pine forests, which occupy about 25-million hectares and annually contribute over 4.5 billion dollars

TABLE 1 Variations in rainfall, temperature, ozone exposure dose and MSI over a May–September interval contributed to the variations in stem growth during the 5-year study

Year	Mean environmental conditions*				Tree growth (mm) [†]	
	Rainfall (mm)	Temperature (°C)	MSI	Ozone (p.p.b.h.)	Wet‡	Dry
1988	4.18	31.1	-0.12 (-1.62)	451	11.54	6.05
1989	7.32	27.7	1.17 (0.19)	162	18.75	13.72
1990	5.69	28.0	0.15 (-0.63)	381	14.89	11.50
1991	5.62	29.0	0.25 (-0.04)	373	13.68	7.51
1992	3.04	26.6	0.53 (-0.02)	306	16.65	11.88

* Daily rainfall (mm) and hourly temperature data (09:00–20:00) collected from the nearby Walker Branch Watershed (2–4 km from the study sites) were provided by the National Oceanic and Atmospheric Administration. Hourly ozone data were collected from a rural station ~30 km from the study site and were provided by the Environmental Protection Agency. Ozone data shown here are average daily sums of concentration per hour ≥ 40 p.p.b. Weekly values of MSI, with seasonal minima in parenthesis, derived for eastern Tennessee were obtained from the National Weather Service. MSI is merely a weekly derivation of the Palmer Drought Severity Index¹⁴, a monthly indicator used for studies of forest growth responses to drought. Values between +1 and -1 are considered normal, whereas values of -2.0, -3.0 or -4.0 represent moderate, severe and extreme drought, respectively, on this log scale index.

† Tree growth rates are shown as means of annual circumference increment per tree for the two drier upland sites (16 trees) and the wetter more fertile stands near a stream bottom (18 trees). The dendrometer bands used in these studies were spring tensioned, marked at two reference points on overlapping ends, and positioned at a height of ~1.5 m. They were measured manually with an electronic calliper with precision on duplicate measurements being typically ± 0.02 mm. Band measurements were initiated on 28 trees in 1988 and 6 trees along the forest edge were added in 1989. The number of measurements per year was 37, 23, 12, 31 and 35 during the 1988–1992 interval. The summed band readings over the 5-year study interval were on average within 0.75 cm of the measured change in circumference using a standard diameter tape. Mean tree diameters ($\pm$ s.d.) were 35.3 ± 4.8 cm at the dry sites and 39.9 ± 4.42 cm at the wet sites.

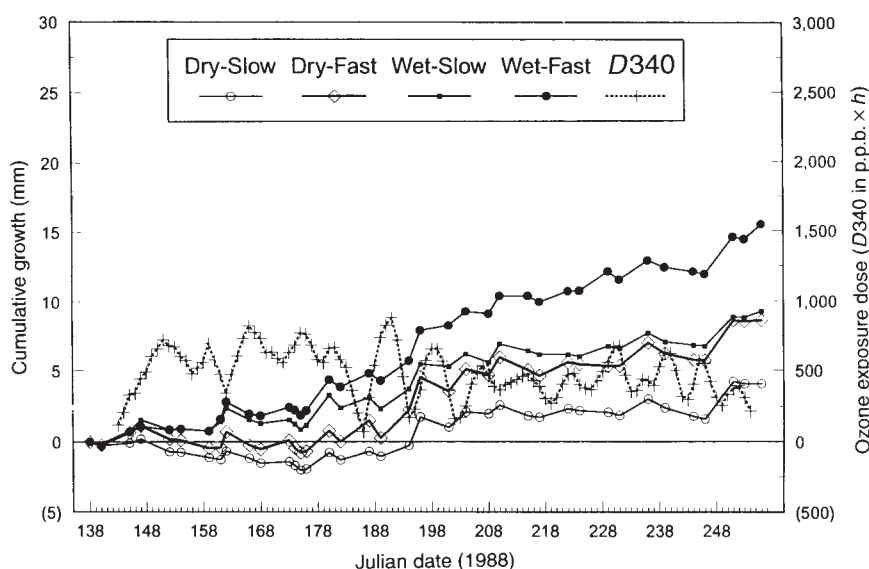
‡ Soil moisture (% volume) varied from 14% to 26% at the drier sites and from 24–38% at the wetter sites for 21 measurements made between 29 May and 6 October 1992. Data were collected at soil depths of 15 cm from each stand using time domain reflectometry¹⁵. Soil analyses representing A-horizon soils of the drier and wetter sites were as follows: total nitrogen (0.089% versus 0.133%), calcium (632 versus 3034 mg kg⁻¹, and P (10.05 versus 11.1 mg kg⁻¹), respectively.

FIG. 1 Variations in seasonal patterns of cumulative circumference growth of mature loblolly pine trees during 1988 demonstrate contrasting growth rates among different sites and subclasses, but common high-frequency fluctuations. Variations among trees are apparent in the mean growth patterns of the three fastest and three slowest growing trees from both the wetter and the drier, less fertile sites. High-frequency patterns were repeated across stands and were closely and inversely related to changing patterns of ozone exposure. Analyses of impacts of ozone and other climatic variables on growth were evaluated by removing the linear time-based growth trends and evaluating partial correlation coefficients¹¹ between high-frequency variations in stem expansion and environmental variables for each individual tree for each year. Ozone or ozone-temperature interactions were significantly ($P < 0.05$) and negatively correlated with stem growth of $\geq 50\%$ of the trees during 3 of 5 years. The 3-day average daily exposure dose at ozone levels $\geq 40 \text{ nl l}^{-1}$ designated D340:

$$D340 = ((\text{sum}_{(\text{day 1 to day 3})} \text{p.p.b. } O_3 \text{ at } \geq 40 \text{ p.p.b.}) \times (H_{0900-2100} \text{ of } O_3 \text{ at } \geq 40 \text{ p.p.b.})) / 3$$

where $H_{0900-2100}$ is the number of hours between 0900 and 2100.

Seasonal variations in D340, shown here, provided the highest statistical correlations with growth responses of the 35 descriptors of ozone exposure examined. These included 6 exposure threshold levels (0, 20, 30, 40, 60 and 80 p.p.b.). Formulations included both concentrations above as well as inclusive of each threshold exposure level. A concentration-weighted function²⁵ was also included. Time intervals included 1, 3 and 7 days preceding the measurement day. Temperature and rainfall were considered at 3- and 7-day intervals. Both ozone and



temperature were also formulated as daily 7-h maximum values as well as in the standard 11-h format for use in 3- and 7-day premeasurement averages. The D340 expression for exposure dose was rated strongest overall for dose description based on its rank among candidate variables in models of growth performance of individual trees and across 5 years of data. Both the 7-day time interval and thresholds of 0, 20 and 30 p.p.b. were good indicators, but not as consistently strong across all trees and years as D340.

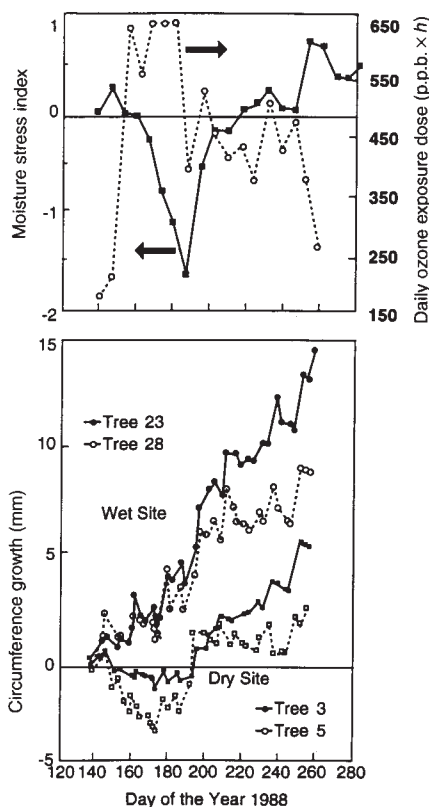


FIG. 2 Comparison plots of the differences in the seasonal dynamics in circumference growth of the fastest- and slowest-growing trees were examined to evaluate the timing and cumulative significance of growth differences of trees in dry and wetter sites during 1988. Note both episodic departures in growth rates between tree categories early in the season followed by chronic differences in growth rate that developed over time. Overlays of the weekly average of 3-day average ozone exposure dose (D340; open circles) and weekly values of moisture stress (MSI; filled squares) are indicated for comparison with emerging trends. Note that high ozone exposure dose developed well in advance of lowest values of MSI. Ozone, temperature and moisture stress are often well correlated with each other in the southeastern United States, because hot, dry years are often associated with air stagnation systems that lead to regional ozone buildup. However, associations of these variables at a weekly time scale may change significantly during a single growing season as shown above. Differences in the timing and relative magnitude of growth responses in relation to changing ozone exposure and MSI as shown in this figure were important in serial analyses of statistical significance of relationships among growth and environmental variables.

to the economy of the southeastern United States⁷. During the mid-1980s, unexplained declines in growth of unmanaged pine stands were detected⁸⁻¹⁰. Air pollution, particularly ozone, has been suggested as a possible cause^{3-5,9}, but verification of growth effects on mature trees has been lacking.

Short-term changes in stem circumference of 28–34 mature trees were monitored at 138 intervals over 5 growing seasons (May–October, 1988–1992) using a sensitive, inexpensive dendrometer band system. The study interval included widely variable temperature, rainfall and ozone exposure conditions and growth rates that varied by 75% across years (Table 1). Sampled trees were canopy dominants (~60 years old) representing widely varying stand density, soil fertility and soil moisture (Table 1).

Comparisons between seasonal growth trends of sampled trees revealed site quality effects (Fig. 1) with superimposed high-frequency variability that was synchronous across sites and closely associated with fluctuations in ozone exposure dose. Episodic responses were sometimes accompanied by cumulative differences in trends apparent only during the last half of the season (Fig. 2). Such changes suggested slower-growing individuals grew even more slowly as the season progressed.

Initial analyses selected environmental variables that were most strongly correlated¹¹ with changes in stem circumference growth (Fig. 2). Both the duration of averaging times and, in the case of ozone, threshold level were considered in selecting the strongest indicator variables. Growth was most consistently influenced by ozone exposures $\geq 40 \text{ nl l}^{-1}$ at 3-day intervals before measurement. Stepwise regression^{12,13} highlighted the variability in environmental influences on growth for each year. Individual trees responded to direct or interactive ozone exposure in direct proportion to the seasonal ozone exposure, ranging from 0 in 1989, the cleanest year, to a maximum of 27 of 28 trees in 1988, the year of highest ozone exposure. Weekly moisture stress index (MSI), based on the relative demands of temperature-driven evapotranspiration compared to available water in the upper soil profile¹⁴, affected a maximum of 26 trees in 1989. Temperature was the dominant influence (34 trees) in 1992, the coolest year.

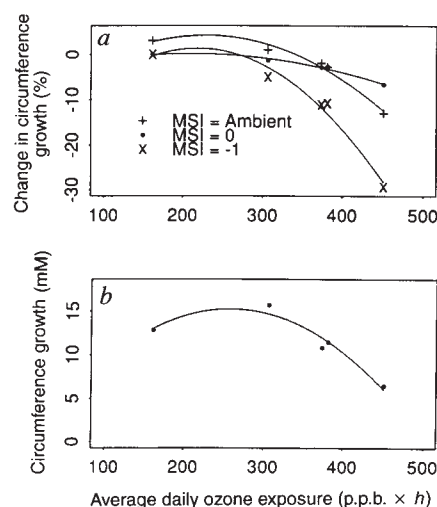
The 5 years of growth data were next combined into an empirical regression model to quantify contributions of environmental variables for the entire data set. The model took the form: $\text{stem growth} = S \times t + 0.00360 \times O \times M + 0.0471 \times R + 0.0458 \times T - 0.0000454 \times O \times T$, where S is the slope of the annual growth trend by year (values for 1988–1992 were 0.0532, 0.117, 0.104, 0.098 and 0.142, respectively); t is the time from the beginning of the annual measurement cycle (about mid-May) in days; O is the 3-day average ozone exposure per day in p.p.b. (where p.p.b.h. is the accumulated hourly dose) for ozone levels $\geq 40 \text{ p.p.b.}$ during the period between 09:00 and 20:00; M is the moisture stress index, calculated weekly; T is the average 3-day air temperature in $^{\circ}\text{C}$ during the hours 09:00 to 20:00; R is the average 3-day rainfall in mm d^{-1} .

Interactions between ozone and moisture stress accounted for 58% of the high-frequency variance in the model, whereas rain, temperature and temperature $\times$ ozone accounted for 25, 12 and 5%, respectively. Model-simulated impacts of ozone on total annual growth, expressed as summed interactions of ozone with both MSI and temperature, ranged from a maximum of 13% in 1988 to near zero in 1989 and 1992. To evaluate the relative significance of MSI and ozone on high-frequency components of growth, simulations were run with variable levels of both ozone and MSI. Results (Fig. 3) suggest that the high-frequency effects of the 300 p.p.b.h. increase in mean daily ozone exposure in the most polluted year (1988) compared to the cleanest year (1989) would reduce stem growth by about 7% in a relatively moist year and by almost 30% in a moderately dry year.

Chronic effects of ozone and interacting environmental variables were suggested by large interannual differences in annual growth trend coefficients noted in the model. The linear growth rates, which include combined effects of all environmental variables, showed a strong inverse relationship to ozone exposure dose (Fig. 3b). At this stage, we cannot precisely define the ozone component of the trend in Fig. 3b. Data in Fig. 3a and additional regression analyses¹⁶ indicate that it is significant relative to other climatic variables.

The rapidity and direction of changes in stem circumference following short-term increases in ozone exposure and the impor-

FIG. 3 a, Ozone effects on growth. The empirical model of stem growth in relationship to rainfall, temperature, MSI and ozone exposure was used to evaluate growth responses to ozone exposure under the ambient environmental regime of each year during the 1988–1992 interval. The model superimposed environmental influences at weekly and shorter scale intervals before each measurement date over 5 years, on the linear growth trends separately determined for each year. Terms included in the model were significant at $\geq 95\%$ confidence level except $T \times D340$ ($P \leq 0.08$). The model explained 99% of the total variation in growth over time. High-frequency variance attributable to physical and chemical climate explained 51% of the variability in the model after seasonal growth trends were removed. Diagnostic criteria²⁶ indicated that multicollinearity among variables was well within acceptable levels. In this figure, model estimates of the impact of ozone under both ambient climatic conditions (plus symbol) and at two levels of soil moisture stress are indicated. The combined effects of differences in temperature and rainfall between the years were calculated to be $\leq 1\%$. The resultant curves (MSI=0 and -1) represent differences in response attributed to increasing the mean daily ozone exposure (D340) from the level experienced during the relatively low pollution year (1989) to levels actually experienced during each year. These estimates represent the effects of high-frequency responses on total seasonal growth. b, Annual growth trends. Relative growth trends for a standard 110-day interval plotted in this figure against annual mean daily ozone exposure dose represent the low frequency components of seasonal growth for each year and were estimated from the linear time trends identified by the empirical model. Differences in both ozone exposure and MSI between years doubtlessly contributed to differences in seasonal growth rate for trees on moist sites. With the inclusion of a sixth year of annual data from 1993¹⁶, 72% of the variation (linear R^2 with $P < 0.02$) in annual growth was accounted for by linear regression of annual stem growth trends against annual variations in ozone exposure. By contrast, seasonal moisture stress index explained 37% ($P < 0.10$) of the variation.



On drier sites contributions of both ozone ($R^2 = 0.40$, $P < 0.10$) and MSI ($R^2 = 0.12$, n.s.) were less significant on an annual scale. Our analyses of the growth patterns of trees on wet versus dry sites¹⁶ and the knowledge that trees on drier sites must allocate relatively more growth into root systems, suggests that trees on wet sites may be more sensitive to amplification of moisture stress by episodic ozone exposure. On the basis of documented seasonal carryover of ozone exposure effects on loblolly pine foliage²⁸, we suspect growth in 1989, which anchors the low ozone end of this curve, may have been influenced by high ozone exposures in the previous year. All curves were fitted using a standard quadratic function.

tant interactions between ozone and MSI (Fig. 3) suggest that alteration of stem water balance may be an important factor in ozone-induced slowing of stem growth. Stem shrinkage occurs in trees under water stress¹⁷ as stored water is moved to areas of greatest water demand. Both altered water uptake through roots and control of water loss through foliage can be inferentially linked to ozone exposure. Reduced root growth has been the most consistent growth response of tree seedlings to ozone in controlled exposure studies⁵. Stomatal closure has frequently been reported with high ozone exposure in controlled studies; however, there is increasing evidence that ozone at near ambient levels can increase water loss of large trees^{18,19} and drought-stressed loblolly pine saplings²⁰. Cuticular weathering of needle surfaces in areas of high wet and dry pollutant deposition^{21,22} may also contribute to decreased control of water loss from pine trees.

The speed of observed changes in stem circumference and their occurrence across widely varying soil water supply in this study suggest ozone-induced amplification of whole-tree moisture stress. With increasing evidence of a plausible mechanism for ozone-induced amplification of transpiration^{18–20}, the three criteria for inferring a cause and effect relationship among associated variables²³ are met: (1) sensitivity of short- and longer-term components of stem circumference growth to changes in ozone exposure level; (2) consistency of observed responses among trees and across sites; and (3) the existence of a plausible mechanism for amplification of moisture stress by ambient ozone exposure¹⁹.

We have described here a research approach for measuring growth responses of mature trees to ozone and climate in an unaltered forest environment. We believe it offers an important research tool for evaluating subtle changes in forest growth to individual and interacting environmental stresses. These analyses have demonstrated both episodic and chronic alteration of stem growth of mature trees associated with ambient levels of ozone. The former we can relate directly to ozone exposure; the latter appears to reflect cumulative, interactive effects of ozone and other climatic stresses. The impact of these effects on regional growth of tree species comparable in sensitivity to loblolly pine

could be important and yet undetected by conventional analyses. The sensitivity of these effects to moisture and temperature may become particularly significant if the climate changes predicted by global climate models²⁴ are realized. Predictions of increases in ozone by 2025 in the large urban–agricultural areas of the world²⁷ further emphasize the potential hazards that these effects pose. □

Received 4 July 1994; accepted 3 February 1995.

1. Finlayson-Pitts, B. J. & Pitts, J. N. Jr *Air Waste* **43**, 1091–1100 (1993).
2. Prinz, B. *Environment* **29**, 10–37 (1987).
3. McLaughlin, S. B. J. *Air Poll. Contr. Assoc.* **35**, 512–534 (1985).
4. Shriner, D. S. et al. *Responses of Vegetation to Atmospheric Deposition and Air Pollution*, Vol. 18 of National Acidic Precipitation Assessment Program (NAPAP) (1990).
5. McLaughlin, S. B. in *Plant Responses to the Gaseous Environment* (eds Alscher, R. G. & Wellburn, A. R. 315–338 (Chapman and Hall, London, 1994).
6. Taylor, G. E. *J. env. Qual.* **23**, 63–82 (1994).
7. United States Department of Agriculture *The South's Fourth Forest: Alternatives for the Future* (U.S. Forest Service Report No. 24.512, 1988).
8. Sheffield, R. M. & Cost, N. D. *J. For.* **87**, 29–33 (1987).
9. Zahner, R. et al. *Can. J. For. Res.* **19**, 612–621 (1989).
10. Bechtold, W. A. et al. *For. Sci.* **37**, 703–717 (1991).
11. Johnson, R. A. & Wichern, D. W. *Applied Multivariate Statistical Analysis* 2nd edn (Prentice Hall, Englewood Cliffs, New Jersey, 1988).
12. SAS/STAT User's Guide Version 6 4th edn N12 (SAS Institute, Cary, NC, USA 1989).
13. Draper, N. R. & Smith, H. *Applied Regression Analysis* 2nd edn (Wiley, New York, 1981).
14. Palmer, W. C. *Meteorological Drought* (US Weather Bureau, Washington DC, 1965).
15. Topp, G. C. & Davis S. L. *Soil Sci. Soc. Am.* **49**, 19–24 (1985).
16. McLaughlin, S. B. & Downing, D. J. *Can. J. For. Res.* (submitted).
17. Kramer, P. J. & Kozlowski, T. T. *Physiology of Woody Plants* (Academic, New York, 1979).
18. Lee, W. S. et al. *Water Air Soil Poll.* **51**, 105–116 (1992).
19. Maier-Markler, U. & Koch, W. *Trees* **7**, 12–25 (1992).
20. Wallin, G. & Skarby, L. *Trees* **6**, 128–136 (1992).
21. Huttunen, S. & Laine, K. *Ann. Bot. Fenn.* **20**, 79 (1983).
22. Fowler, D., Cape, J. N., Nicholson, I. A., Kinnaird, J. W. & Patterson, I. S. in *Proc. Int. Conf. Ecol. Impact of Acid Precip.* 146 (eds Drablos, D. & Tolan, A.) (SNSF, Norway, 1980).
23. Mosteller, F. & Tukey, J. *Data Analysis and Regression: A Second Course in Statistics* (Addison-Wesley, Reading, MA, 1977).
24. Joyce, L. A. et al. *Climate Change and America's Forests* (USDA Gen. Tech. Rep. RM-187, 1990).
25. LeFohn, A. S. et al. *Atm. Env.* **26**, 287–298 (1992).
26. Belsky, D. A., Kuh, E. & Welsch, R. E. *Regression Diagnostics* (Wiley, New York, 1980).
27. Chameides, W. L., Kasibhata, P. S., Yienger, J. & Levy, Y. *Science* **264**, 74–77 (1994).
28. Sesak, T. W. et al. *For. Sci.* **37**, 1078–1098 (1991).

ACKNOWLEDGEMENTS. We thank J. D. Joslin, W. K. Roy, A. Stone and B. L. Jackson for data collection and compilation and D. Marx (US Forest Service), R. J. Norby, T. J. Blasing and D. S. Shriner for advanced technical review of this manuscript. This research was sponsored by the US Forest Service Global Change Research Program through Interagency Agreement with the US Department of Energy.

Primary production required to sustain global fisheries

D. Pauly* & V. Christensen

International Center for Living Aquatic Resources Management, MCPO Box 2631, 0718 Makati, Metro Manila, Philippines

* Present address: Fisheries Centre, University of British Columbia, 2204 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada.

THE mean of reported annual world fisheries catches for 1988–1991 (94.3 million t) was split into 39 species groups, to which fractional trophic levels, ranging from 1.0 (edible algae) to 4.2 (tunas), were assigned, based on 48 published trophic models, providing a global coverage of six major aquatic ecosystem types. The primary production required to sustain each group of species was then computed based on a mean energy transfer efficiency between trophic levels of 10%, a value that was re-estimated rather than assumed. The primary production required to sustain the reported catches, plus 27 million t of discarded bycatch, amounted to 8.0% of global aquatic primary production, nearly four times the previous estimate. By ecosystem type, the requirements were only 2% for open ocean systems, but ranged from 24 to 35% in fresh water, upwelling and shelf systems, justifying current concerns for sustainability and biodiversity.

Global primary productivity generates annually about 224×10^9 t dry weight of biomass. Of this, 59% is produced in terrestrial ecosystems, the rest in aquatic systems¹. Of the terrestrial primary production, 35–40% is presently used by humans, directly (for example, as food or fibre), indirectly (for example, as feed for animals) or foregone (through, for example, urban sprawl)¹. This was estimated by adding the primary production required (PPR) by various production systems (such as cultivated and grazing lands) or by various subsectors (such as timber or fibre production), thus allowing errors in the independent subtotals to cancel out partly.

The PPR to sustain the world's catches of 75 million t in the early 1980s was also estimated in the same study, based on the key assumption that the 'average fish' feeds two trophic levels above the primary producers. This suggested that 2.2% of the world's aquatic primary production was required to sustain the fisheries, and thus led to the conclusion that 'human influence on the lowest trophic levels in the ocean (outside severely polluted areas) is minimal, and human exploitation of marine

resources therefore seems insufficient by itself to alter on a large scale any but the target populations and those of other species interacting closely with target species'¹.

This work is an attempt to obtain a more accurate estimate of the PPR to sustain the world fisheries catches (including discarded 'bycatch'), based on the same approach as used above to estimate terrestrial PPR, wherein independent estimates are obtained on a commodity group and system basis, then added up to yield a robust estimate of the total.

Our approach, illustrated in Fig. 1, uses only flows of matter (catches and food consumption of fishes and their prey) and does not require estimation of biomasses, which have proved hard to estimate reliably on a global basis².

Recent world fisheries statistics, covering a short period (1988–1991) without major changes in catch composition and reported by the UN Food and Agriculture Organization (FAO)³, were split into 39 commodity groups, by ecosystem types. The PPR was then estimated by group, and ecosystem type, based on an estimate of 10% mean transfer efficiency between trophic levels (Fig. 2), and the mean trophic levels of the commodity groups (Table 1). This led to group-specific estimates of PPR, presented here by ecosystem type, after accounting for the 27 million t of discarded bycatch recently estimated, based on thorough review of worldwide discarding practices⁴ (Table 2).

The results differ markedly from those of the previous study based on an 'average fish': we estimate that 8.0% of the world's aquatic primary production is required to sustain the fisheries, nearly four times the earlier estimate. The difference is due to our use of higher fisheries catches, our consideration of discards, and the fact that we used disaggregated data, to account for the non-linearity of the relationship between PPR and trophic levels.

Although our 8% still may be a moderate figure compared to 35–40% for the terrestrial systems, the prospects for increases are dim. The bulk of aquatic productivity (75%) occurs in the open ocean (gyre) systems, by virtue of their vast extent. Only 1.8% of this productivity is used, but as little as 20 to 25% of the overall zooplankton biomass may be available for the higher trophic levels⁵, dominated by top predators (notably yellowfin and skipjack tuna), which must roam desert-like ocean expanses to find scattered food patches.

The estimated PPR for the coastal and coral reef systems is 8.3%. This relatively low value is due to (1) a high level of productivity (Table 2), (2) large catches at low trophic levels (seaweeds, bivalves and other invertebrates), and (3) overfishing,

FIG. 1 Schematic representation of approach used here to estimate the PPR to sustain the catches of a given ecosystem. Left, the simplest among the 48 trophic models used here, representing the lightly fished Lake Turkana¹⁸, each of its state variables (boxes) has inputs (food) and outputs (predation and/or fishery catches), in t wet wt km⁻² yr⁻¹; only major flows are shown, excluding respiration and back flows to the detritus. Right, the pyramid illustrates how the catches (circles) are raised to PPR at trophic level 1 (diamonds), using the 10% trophic efficiency of Fig. 2.

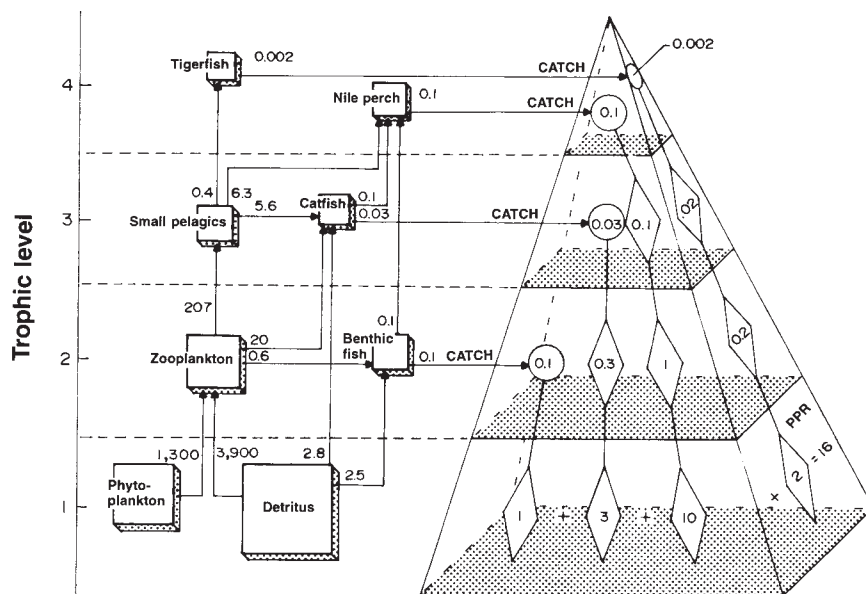


TABLE 1 Reported world fisheries catches, ancillary statistics and the PPR to sustain these catches

FAO-codes	Species group	Catch (ww; t × 10 ³)	n	k	Trophic level Mean	s.e.	PPR (g C × 10 ¹²)
Oceanic (gyre) systems							
36	Tunas, bonitos, billfishes	2,975	1	3	4.2	0.04	523.9
46	Krill	344	—	—	2.2*	—	0.6
Upwelling systems							
35	Anchovies, sardines	11,597	24	97	2.6	0.28	53.1
34	Jacks	4,785	8	28	3.2	0.06	86.7
37	Mackerels	1,096	10	44	3.3	0.10	22.8
57	Squids†	248	6	31	3.2	0.14	6.9
Tropical shelves							
24, 35	Small pelagics	7,127	5	20	2.8	0.27	59.9
31, 33, 39	Misc. teleosteans	5,342	22	16	3.5	0.26	204.3
34, 37	Jacks, mackerels	2,053	8	46	3.3	0.28	45.5
36	Tunas, bonitos, billfishes	1,275	8	44	4.0	0.12	141.7
57	Squids, cuttlefishes, octopuses	1,114	6	31	3.2	0.14	19.6
45	Shrimps, prawns	650	4	21	2.7	0.35	35.0
42–44, 47, 77	Lobster, crabs and other invertebrates	544	7	35	2.6	0.30	2.2
38	Sharks, rays, chimaeras	344	9	51	3.6	0.24	15.2
Non-tropical shelves							
32	Cods, hakes, haddocks	12,209	5	49	3.8	0.25	929.9
33	Redfishes, basses, congers	3,837	2	5	3.4	0.06	110.9
39	Miscellaneous marine fishes	3,362	1	5	3.2	0.11	52.8
34	Jacks, mullets, sauries	2,871	1	3	3.8	0.13	206.0
35	Herrings, sardines, anchovies	2,319	3	8	3.0	0.15	23.7
42–45, 47, 75, 77	Shrimps and other crustaceans	1,195	3	10	2.3	0.24	2.6
57	Squids, cuttlefishes, octopuses†	1,114	6	31	3.2	0.14	19.3
31	Flounders, halibuts, soles	1,098	3	10	2.9	0.12	9.8
37	Mackerels, cutlassfishes	1,096	3	16	3.4	0.29	30.6
23–25	Diadromous fishes	819	14	49	2.4	0.25	2.3
38	Sharks, rays, chimaeras	344	2	15	3.7	0.28	19.2
Coastal and coral systems							
52–56, 58	Bivalves and other molluscs	5,150	4	12	2.1	0.13	7.6
31, 39	Miscellaneous marine fishes	3,424	15	86	2.8	0.41	24.0
35	Herrings, sardines, anchovies	2,319	9	52	3.2	0.20	40.8
9	Seaweeds	1,683	1	—	1.0	—	0.2
34, 37	Jacks and mackerels	1,322	17	97	3.3	0.22	29.3
23–25	Diadromous fishes†	819	3	13	2.8	0.19	5.7
43–45, 47	Shrimps, prawns	748	8	42	2.6	0.33	3.3
42, 74–77	Crustaceans and other invertebrates	566	14	49	2.4	0.25	1.6
72	Turtles	2	2	7	2.4	0.37	0.006
Freshwater systems							
13	Misc. freshwater fishes	5,237	41	273	3.1	0.28	69.4
21–25	Misc. diadromous fishes	1,210	23	121	3.6	0.27	60.1
41, 45, 51, 54, 71, 77	Invertebrates and amphibians	896	14	54	2.2	0.23	1.6
11	Carp-like fish	632	15	79	2.7	0.34	3.7
12	Tilapias and other cichlids	579	24	11	2.5	0.18	2.0

The items used to infer primary production required (PPR) from the reported annual catches (wet weight; means for 1988–1991) were: (1) codes to link the groups in the FAO catch statistics³ to those in 48 trophic models of ecosystems whose sources are given in Fig. 2; (2) major group names; (3) group catch (total average catches less aquaculture production for 1988 of freshwater, brackishwater and marine fish fed on artificial fish farms, except for India whose annual production of carp-like fishes was assumed at 2×10^5 t), assigned to ecosystem type based on ecological and geographic considerations; (4) mean trophic level (TL), estimated for the groups included in the models in Fig. 2, with standard errors (s.e., in brackets) estimated from $\sum_{i=1}^k s_i^2(n_i - 1)/(n - k)$ where n_i is the number of prey items used for one estimate of TL, n the sum of all n_i , k the number of TL estimates used to compute each group's mean TL, and the s_i^2 are the variances of these TL estimates, that is, the indices of omnivory¹¹; (5) values of n_i and k . The PPR estimates are based on a conservative 9:1 ratio for the conversion of wet weight to carbon¹² and a 10% transfer efficiency per trophic level (Fig. 2), that is, using $PPR = (\text{catches}/9) \times 10^{(TL - 1)}$.

* Assumed diet composition 80% phytoplankton, 20% herbivorous zooplankton¹⁰.

† No system specific estimate of trophic level available; instead a value from another system type is used.

which has left the reduced fish biomass unable to use the available production.

The shelf systems exhibit high PPR, from 24.2 to 35.3%, mainly due to industrialized fisheries operating at high trophic levels, a feature in which they differ from upwelling systems, for which, however, the PPR is still high, 25.1% (Table 2). It would seem difficult to further increase these values, especially on temperate shelves, given that a substantial part of the primary production generated during intensive blooms settles out as detritus, before use by zooplankton is possible. Also, higher PPR would starve top predators, such as marine mammals and birds.

Could catches be increased through fishing down the food web, that is, concentrating on fishes at lower trophic levels? In

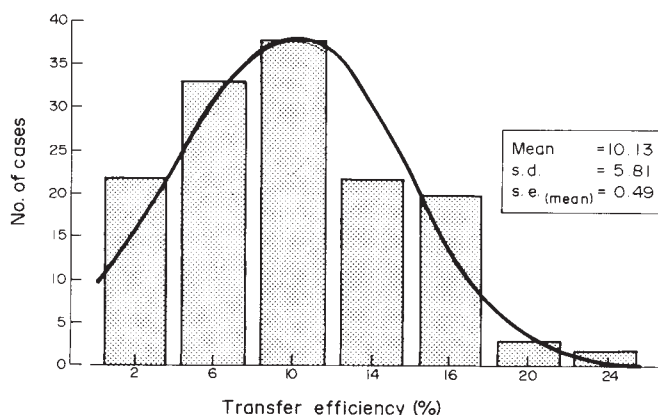
ocean systems, the fisheries targeting tuna would have to harvest the prey of tuna (mainly small pelagic fish), which cannot be done economically at present. In coastal and coral reef systems, fishing has already moved down the food web, and improvements must come from rebuilding biomasses through better management^{6,7}. This is also true for tropical shelves, where intensive overfishing is causing significant loss of spawning biomass and of biodiversity, especially through shrimp trawling on soft-bottoms, which results in destruction of soft corals and massive changes in community structure, including large fish being replaced by short-lived organisms (small pelagic fish, cephalopods, jellyfish and so on)⁷. This is aggravated by the fact that, contrary to some terrestrial ecosystems such as rainforests,

TABLE 2 Global estimates of primary production (PP), of PPR to sustain world fisheries (mean for 1988–1991, wet weight), and of the mean trophic levels (TL) of the catches, by ecosystem type

Ecosystem type	Area (10 ⁶ km ²)	PP (gC m ⁻² yr ⁻¹)	Catch (g m ⁻² yr ⁻¹)	Discards (g m ⁻² yr ⁻¹)	TL of catch	PPR (catches + discards)	
						Mean (%)	95% Confidence interval
Open ocean	332.0	103	0.01	0.002	4.0	1.8	1.3–2.7
Upwellings	0.8	973	22.2	3.36	2.8	25.1	17.8–47.9
Tropical shelves	8.6	310	2.2	0.671	3.3	24.2	16.1–48.8
Non-tropical shelves	18.4	310	1.6	0.706	3.5	35.3	19.2–85.5
Coastal/reef systems	2.0	890	8.0	2.51	2.5	8.3	5.4–19.8
Rivers and lakes	2.0	290	4.3	n.a.	3.0	23.6	11.3–62.9
Weighted means (or total)	(363.8)	126	0.26	0.07	2.8	8.0	6.3–14.4

Distribution of surface areas by ecosystem type was estimated based on planimetry, checked against published estimates¹³. As defined here, coastal systems generally reach down to a depth of 10 m, except coral reefs which may reach to about 30 m¹⁴. The PP estimates are based on ref. 15, with a 1:3 allocation between ocean and shelf productivity¹², and using newer, higher values for upwelling systems¹⁶, leading to a more conservative estimate of PPR. The nutrients released by discarded fish are assumed to have negligible effects on primary production, and on food webs in general¹⁷. The catches are from Table 1 and their TL are weighted means. The PPR estimates are from Table 1; they were adjusted to account for the discards (allotted across systems based on group-specific catch/discard ratios⁴). Their standard errors (s.e.) were estimated by the Monte Carlo method assuming normal distributions, with the s.e. of the mean transfer efficiencies in Fig. 2, and of the mean TL in Table 1 providing all the variability (10,000 runs per group in Table 1, all within ± 1 s.e. of each mean), the FAO catches³, the discards⁴, and the PP by ecosystem type being used as fixed values, notwithstanding their imprecision (see text).

FIG. 2 Frequency distribution of energy transfer efficiencies (TE, in %) in 48 trophic models of aquatic ecosystems. The estimates of TE ($N = 140$) express, for TL 2 (=herbivores and detritivores) to 4 (=third-order consumers), the fraction of production passing from one TL to the next, and account for consumers feeding on the different TL of an ecosystem^{11,19} (no trend of TE with TL was apparent²⁰). All 48 trophic models used as sources of TE are fully documented^{11,16,21–23}; they jointly show that the 10% TE value commonly used for aquatic organisms²⁴ is extremely close to the mean of the 140 estimates available for this study.



of which large undisturbed tracts still exist, and contrary to what is stated in the introductory quote, the overwhelming bulk of the world's trawable shelves is impacted by fishing, leaving few sanctuaries where biomasses and biodiversity remain high.

There is at present a debate about the level of global primary production, which may be slightly higher than the figure we have used^{8,9}. However, our fisheries catches are likely to be underestimated as well, despite having been adjusted for discarding practices, because of under-reporting and under-collection, assumed by numerous fisheries scientists, but still awaiting the kind of global analysis now done for discards⁴.

Our results, having been obtained by summing independent group- and system-specific catches and PPR estimates, should be robust. Moreover, the estimate of PPR of 8% for all aquatic systems, although nearly four times as high as the previous estimate, masks the important fact that the nearshore systems readily accessible to humans (most upwellings and the shelves), and the freshwater systems have very high PPR, nearing that estimated for terrestrial systems. Because we do not have the ability safely to increase aquatic primary production, these high PPR values confirm broad limits on the carrying capacity of natural aquatic ecosystems, which still form the basis for 85% of the world fish harvest. □

Received 24 August; accepted 26 January 1995.

1. Vitousek, P. M., Ehrlich, P. R., Enrich, A. H. & Matson, P. A. *Bioscience* **36**, 368–373 (1986).
2. Gulland, J. A. (ed.) *The Fish Resources of the Oceans* (Fishing News Books, West Byfleet, UK, 1971).
3. FAO *FAO Year book* **72**, 113–120 (1993).

4. Alverson, D. L., Freeberg, M. H., Murawski, S. A. & Pope, J. G. *FAO Tech. Pap.* 339 (1994).
5. Roger, C. *Envir. Biol. Fishes* **39**, 161–172 (1994).
6. May, R. M., Beddington, J. R., Clark, C. W., Holt, S. J. & Laws, R. M. *Science* **203**, 267–277 (1979).
7. Pauly, D. in *Ocean Yearbook 6* (eds Borgese, E. M. & Ginsburg, N.) 29–37 (Univ. Chicago Press, Chicago, 1986).
8. Li, W. K. W. et al. *Science* **219**, 292–295 (1983).
9. Post, W. M. et al. in *The Science of Global Change* (eds Dunnette, D. A. & O'Brien, R. J.) 392–412 (Am. Chem. Soc. Symp. Ser. Washington DC, 1992).
10. McClatchie, S. *Continental Shelf Res.* **8**, 329–345 (1988).
11. Christensen, V. & Pauly, D. *Ecol. Modelling* **61**, 169–185 (1992).
12. Strathmann, R. R. *Limnol. Oceanogr.* **12**, 411–418 (1967).
13. Lieth, H. in *Patterns of Primary Production in the Biosphere* (ed Lieth, H. F.) 277–282 (Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1978).
14. Crossland, C. J., Hatcher, B. G. & Smith, S. V. *Coral Reefs* **10**, 55–64 (1991).
15. De Voors, G. G. N. in *The Global Carbon Cycle* (eds Bolin, B., Degens, E. T., Kempe, S. & Ketner, P.) 259–292 (Wiley, New York, 1979).
16. Jarre-Teichmann, A. & Christensen, V. *International Centre for Living Aquatic Resources Management Studies and Reviews* **24** (in the press).
17. Cushing, D. H. in *Penaeid Shrimps: their Biology and Management* (eds Gulland, J. & Rothschild, B.) 254–258 (Fishing News Books, Farnham, Surrey, England, 1984).
18. Kolding, J. in *Trophic Models of Aquatic Ecosystems* (eds Christensen, V. & Pauly, D.) 116–123 (International for Living Aquatic Resources Management, Manila, 1993).
19. Cousins, S. *New Scientist* **4**, 50–54 (1985).
20. Christensen, V. & Pauly, D. (eds) in *Trophic Models of Aquatic Ecosystems* 338–352 (International Center for Living Aquatic Resources Management, Manila, 1993).
21. Christensen, V. *International Council for the Exploration of the Sea, Council Meeting/L:25* (1992).
22. Christensen, V. & Pauly, D. (eds) *Trophic Models of Aquatic Ecosystems* (International Center for Living Aquatic Resources Management, Manila, 1993).
23. Pauly, D. & Christensen, V. in *Large Marine Ecosystems: Stress, Mitigation and Sustainability* (eds Sherman, K., Alexander, L. M. & Gold, B. D.) 148–174 (American Association for the Advancement of Science, Washington DC, 1993).
24. May, R. M. in *Theoretical Ecology: Principles and Applications* (ed. May, R. H.) 142–162 (Blackwell Scientific, Oxford, 1976).

ACKNOWLEDGEMENTS. We thank R. E. Ulanowicz for suggesting an approach to aggregate flows in ecosystem models by trophic levels; T. Platt and F. Chavez for advice on global primary productivity; the nearly 100 authors of the ECOPATH II models used here, for their cooperation; and the Danish International Development Assistance for partial funding of the work which led to this contribution.

Aberrant differentiation of neuromuscular junctions in mice lacking s-laminin/laminin $\beta 2$

Peter G. Noakes^{*†}, Medha Gautam[‡],
Jacqueline Mudd[‡], Joshua R. Sanes^{*}
& John P. Merlie^{‡§}

Departments of ^{*}Anatomy and Neurobiology,
[‡]Molecular Biology and Pharmacology,
Washington University Medical School, 660 S. Euclid Avenue,
St Louis, Missouri 63110, USA

SYNAPSE formation requires a complex interchange of information between the pre- and postsynaptic partners. At the skeletal neuromuscular junction, some of this information is contained in the basal lamina (BL), which runs through the synaptic cleft between the motor nerve terminal and the muscle fibre. During regeneration following injury, components of synaptic BL can trigger several features of postsynaptic differentiation in the absence of the nerve terminal, and of presynaptic differentiation in the absence of the muscle fibre¹⁻³. One nerve-derived component of synaptic BL, agrin, is known to affect postsynaptic differentiation³, but no muscle-derived components have yet been shown to influence motor nerve terminals. A candidate for such a role is s-laminin (also called laminin $\beta 2$), a homologue of the B1 ($\beta 1$) chain of the widely distributed BL glycoprotein, laminin³⁰. s-Laminin is synthesized by muscle cells⁵ and concentrated in synaptic BL⁴. *In vitro*, recombinant s-laminin fragments are selectively adhesive for motor neuron-like cells, inhibit neurite outgrowth promoted by other matrix molecules, and act as a 'stop signal' for growing neurites^{6,7}. By generating and characterizing mice with a targeted mutation of the s-laminin gene, we show here that s-laminin regulates formation of motor nerve terminals.

A 10-kilobase (kb) fragment of the mouse s-laminin gene (M.G., P.G.N., J.R.S. and J.P.M., manuscript in preparation) was mutated by inserting a selectable marker into the second of 32 coding exons (Fig. 1a), and the mutation was transferred to embryonic stem (ES) cells by homologous recombination (Fig. 1b). A recombinant clone was injected into blastocytes to generate germline chimaeras. Heterozygotes were apparently normal, and the number of homozygotes born was as predicted (68/325 or 21%), indicating that prenatal development can proceed without s-laminin. Mutants ($-/-$) were externally indistinguishable from their littermates for several days postnatally. However, they stopped growing at about 1 week old, became progressively weaker, and died between postnatal days (P) 15 and 30. We have so far found two abnormalities that could account for the mutant's failure to thrive. First, mutants develop proteinuria, indicating a functional defect in filtration by the renal glomerulus. This defect is likely to result from alterations in the glomerular BL, which is rich in s-laminin⁴. These renal defects, which resemble those seen in a human congenital disease called minimal change nephrotic syndrome, will be described elsewhere (P.G.N., J. Miner, M.G., J.R.S. and J.P.M., submitted). Second, we observed defects in the structure, function and molecular architecture of the neuromuscular junction. Here we will focus on the neuromuscular phenotype.

To initiate studies of neuromuscular junctions, we stained whole P15 muscles with rhodamine- α -bungarotoxin to label acetylcholine receptors (AChRs) in the postsynaptic membrane. Most, if not all, muscle fibres bore a single synaptic site, and

these were clustered near the nerve trunk in the centre of the muscle⁸, forming a normal endplate 'band' (Fig. 1c, d). A combination of cholinesterase⁹ and silver staining¹⁰ showed that each endplate was innervated by the terminal branches of a single motor axon (Fig. 1g, h). Most mutant endplates were innervated by a single axon, suggesting that synapse elimination, which removes polyneuronal innervation from normal endplates during the first two postnatal weeks¹¹, occurred on schedule in the mutant.

Although their numbers and arrangements were normal, mutant endplates were abnormal in shape. In wild-type neonates, endplates are oval-shaped plaques which gradually ramify to form branched (pretzel-shaped) structures in which pre- and postsynaptic specializations are closely apposed (Fig. 1f, j). In the mutant, in contrast, both nerve terminals (stained with a lipophilic anterograde tracer¹²) and postsynaptic membranes (stained with rhodamine- α -bungarotoxin) remained relatively unbranched (Fig. 2e, i).

Such defects might reflect general retardation of maturation in the $-/-$ animals, but electron microscopy revealed that neuromuscular structure was aberrant in at least four specific ways (Fig. 2 and Table 1). First, active zones (focal juxtamembranous densities at which synaptic vesicles fuse with the nerve terminal membrane to release neurotransmitter) were common in normal terminals but rarely seen in mutant terminals. Second, whereas about two-thirds of the synaptic vesicles are normally concentrated in the half of the terminal that abuts the muscle fibres (ref. 1 and Table 1), vesicles were evenly distributed throughout the terminals of s-laminin-deficient mice. This abnormal arrangement reflects a redistribution rather than simply a loss of vesicles from regions near active zones, because the total number of vesicles in mutant terminals was nearly normal.

TABLE 1 Ultrastructural analysis of motor nerve terminals in s-laminin mutant mice

Age	P8		P15	
Genotype (number of endplates)	Control (20)	Mutant (44)	Control (14)	Mutant (30)
Nerve-muscle apposition				
Length (μ m)	4.7 $\pm$ 0.4	4.8 $\pm$ 0.5	4.1 $\pm$ 0.7	3.2 $\pm$ 0.3
% N-M	98%*	22%†	98%*	35%
% N-S-M	2%*	75%	2%*	65%
Junctional folds				
Number/terminal profile	6.4 $\pm$ 1.2*	1.0 $\pm$ 0.2	6.9 $\pm$ 1.0*	0.7 $\pm$ 0.2
% N-M	100%	10%	100%	14%
% N-S-M	0%	90%	0%	86%
Active zones				
Number/terminal profile	1.9 $\pm$ 0.3*	0.2 $\pm$ 0.1	0.9 $\pm$ 0.3*	0.03 $\pm$ 0.03
Per cent of profiles with ≥1 active zone	75%	18%	63%	3%
Mitochondria				
Number/terminal profile	7.7 $\pm$ 1.4	9.2 $\pm$ 1.7	7.0 $\pm$ 2.0	5.7 $\pm$ 1.2
Synaptic vesicles				
Number/terminal profile	162 $\pm$ 16	159 $\pm$ 20	ND	ND
Per cent in juxtamuscular half	64.3 $\pm$ 2.8*	51.7 $\pm$ 1.1	70.4 $\pm$ 4.0*	53.1 $\pm$ 2.3

Electron micrographs were taken at an original magnification of $\times 10,000$; see Fig. 2 for examples and legend for methods. Sections were taken at intervals of $\geq 20 \mu$ m to avoid double-counting individual endplates. Measurements are included from sternomastoid, intercostals and diaphragm. Results did not differ significantly among muscles, so data have been pooled. Similarly, no differences were detected between $+/-$ and $+/+$ endplates, so data from both are combined ('control'). Length of nerve-muscle apposition was measured as for Fig. 2d and e. The total contact area was then divided into portions in which the nerve terminal directly apposed muscle fibre BL (N-M) and those in which a Schwann cell process was present between nerve and muscle membranes (N-S-M). Counts of junctional folds, active zones, vesicles, and mitochondria in controls and mutants can be compared directly at each age, because control and mutant synaptic profiles were similar in size.

* Difference between mutants and littermate controls significant at $P < 0.01$ by Student's *t*-test.

† At an additional 3% of contact area, the nerve terminal was directly apposed to the muscle fibre membrane, with no intervening BL.

[†] Present address: School of Anatomy, University of New South Wales, Sydney, NSW 2033, Australia.

[§] To whom correspondence should be addressed.

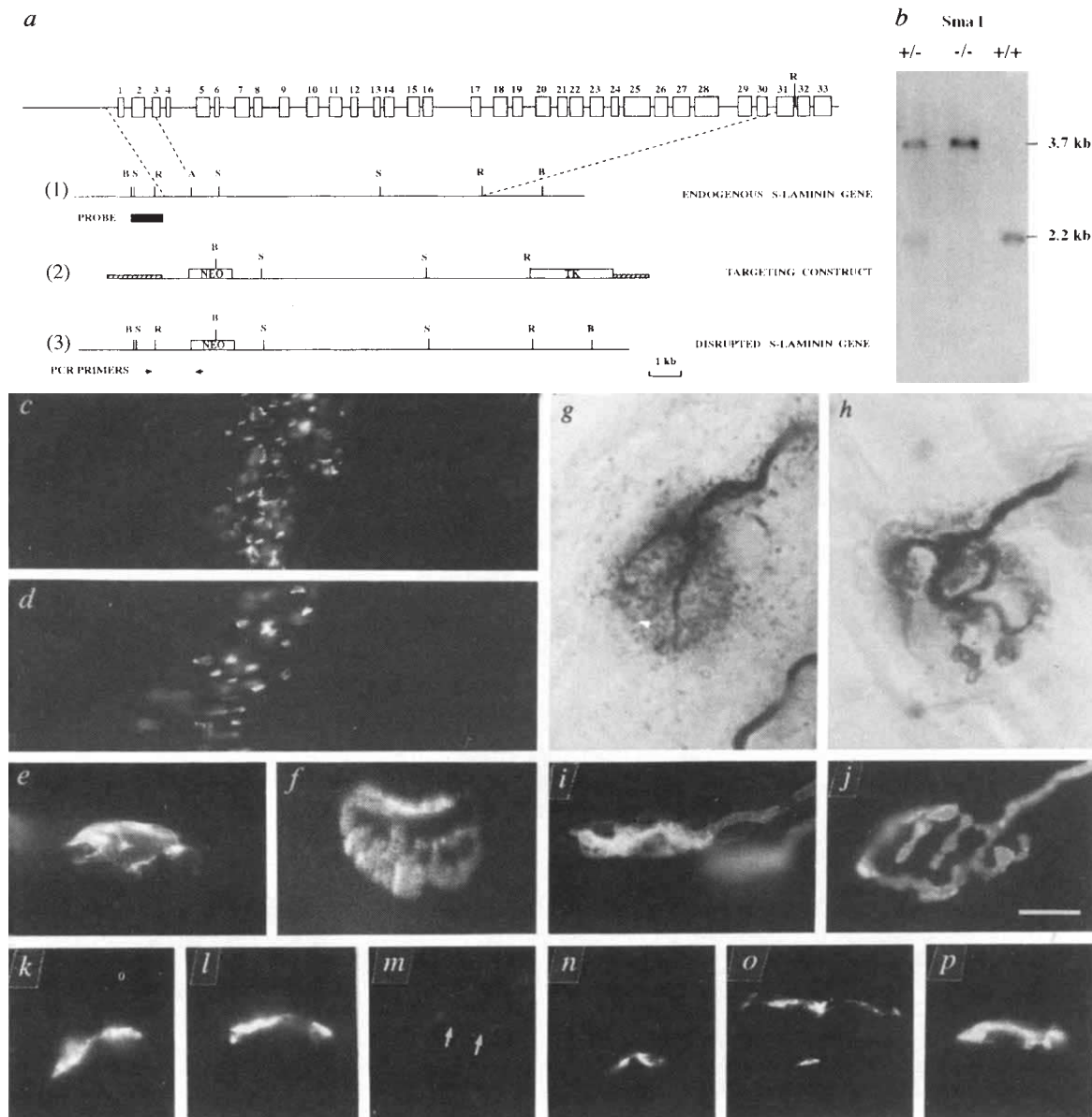


FIG. 1 Light-microscopic analysis of neuromuscular junctions in s-laminin-deficient mutant mice. **a**, The intron-exon structure of the murine s-laminin gene (LAMB2), (top), an s-laminin genomic clone (1), the targeting vector constructed from it (2), and the predicted product of the homologous recombination event (3). **b**, Southern blot analysis of the three genotype categories derived from heterozygote crosses. As predicted from the genomic structure (**a**), the 2.2-kb *Sma*I band in the endogenous gene is replaced by a 3.7-kb band in the mutant. Immunoblotting confirmed the absence of s-laminin protein (not shown). **c** and **d**, Distribution of neuromuscular junctions in s-laminin mutant mice (**c**) and littermate controls (**d**). P15 diaphragms were stained with rhodamine- α -bungarotoxin to identify synaptic sites. In both muscles, endplates are concentrated in a central band, midway between rib and central tendon. Single bungarotoxin-stained endplates from the muscles in **c** and **d** are shown at higher magnification in **e** (mutant) and **f** (control). **g**–**j**, Nerve terminals from P15 muscles stained by a silver method (**g**, mutant; **h**, control) or labelled with Dil (1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate) (**i**, mutant; **j**, control). Mutant nerve terminals and postsynaptic membranes are less branched than controls. **k**–**p**, Immunohistochemical analysis of nerve terminals in intercostal muscles of P15 s-laminin mutant mice (**k**, **m** and **o**) and littermate controls (**l**, **n**, **p**). Sections were stained with antibodies against either synaptotagmin (**k**, **l**) or synapsin (**m**–**p**). Following incubation with antibody, all sections were doubly stained with fluorescein-labelled second antibody to reveal the primary antibody, and with rhodamine- α -bungarotoxin to mark synaptic sites. Position of endplate

is marked with arrows in **m**. Sections in **o** and **p** were fixed with formaldehyde before staining; other sections were unfixed. Mutant endplates bear almost normal levels of synaptotagmin but are deficient in synapsin. Bar in **j** is 100 μ m for **c** and **d** and 10 μ m for **e**–**h** and **k**–**p**.

METHODS. **a** and **b**, To generate the targeting vector, MC1neo-pA- β -actin, a cassette containing the β -actin enhancer and the *neo* gene²⁸ was inserted at the indicated *Agel* site of the *NotI*–*EcoRI* genomic fragment subcloned into Bluescript IISK+ (Stratagene). Arrowheads indicate primers used for PCR screening, and a bar shows the probe used to confirm targeting. Restriction sites used in the construction of the vectors and analysis of the mutants are: A, *Agel*; B, *Bss*HI; R, *EcoRI*; S, *SmaI*. The *Agel* site is not unique in the gene. The probe for Southern analysis was a 440-bp *Bss*HI–*Sau*3A genomic fragment not contained in the targeting vector. The targeting vector contains stop codons in all three reading frames, and a frameshift would occur if the disrupted exon were skipped. The vector was transferred to R1 ES cells²⁹ by electroporation. Approximately 1% of G418-resistant clones were homologous recombinants. **c**–**f**, Diaphragms were fixed in 2% paraformaldehyde, then incubated overnight at 4 °C in rhodamine- α -bungarotoxin. **g**, **h**, Silver-cholinesterase staining was as described in ref. 10. **i**, **j**, Dil-impregnated plugs were applied to the nerves of formaldehyde-fixed sternomastoid muscles, and preparations were incubated at 37 °C for 1 month¹². **k**–**p**, Unfixed muscles were frozen in liquid nitrogen-cooled isopentane, cross-sectioned at 6 μ m in a cryostat, and doubly stained with antibody and rhodamine- α -bungarotoxin as described^{4,14}. Sources of antibodies are given in Table 2.

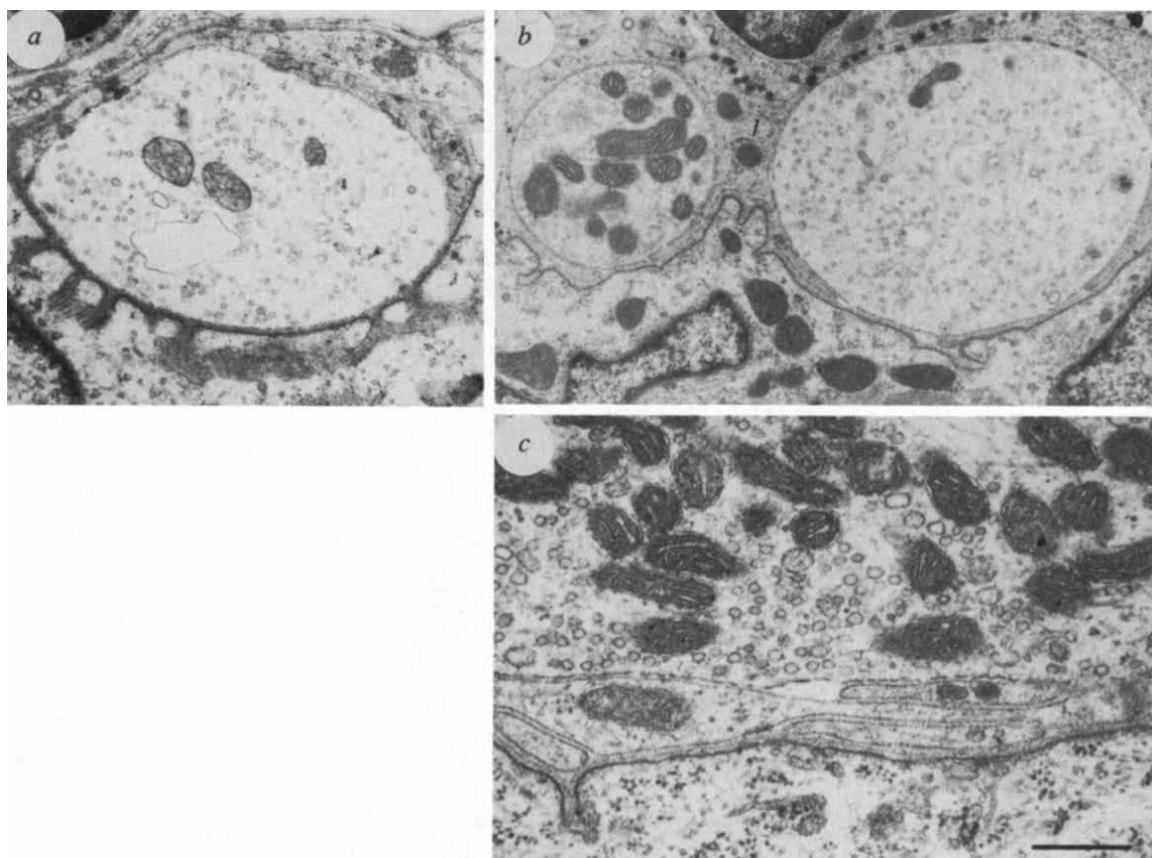
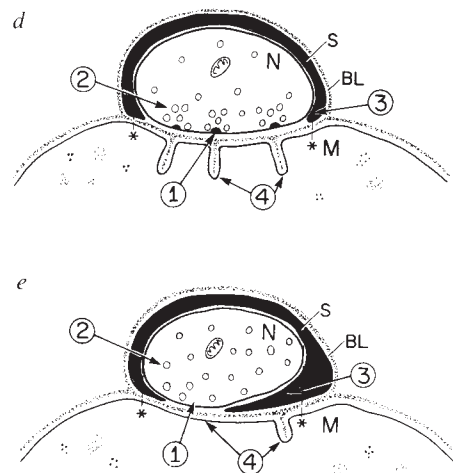


FIG. 2 Ultrastructure of neuromuscular junctions in s-laminin mutant mice (*b*, *c* and *e*) and littermate controls (*a* and *d*). Micrographs (*a*–*c*) are from P8 sternomastoid (neck) muscles, but similar results were obtained from intercostals and diaphragm (Table 1). *d* and *e*, Sketches showing the main ultrastructural effects of s-laminin deficiency: (1) lack of active zones; (2) dispersal of synaptic vesicles; (3) intrusion of Schwann cell processes into the synaptic cleft; and (4) paucity of junctional folds. N, nerve terminal; M, muscle fibre; S, Schwann cell. Asterisks indicate how length of nerve–muscle apposition was measured in Table 1. Bar is 1 μ m for *a* and *b*, 0.5 μ m for *c*.

METHODS. Muscles were fixed in 4% glutaraldehyde plus 4% paraformaldehyde in phosphate buffer, washed thoroughly in buffer, then refixed in 1% OsO₄, dehydrated, and embedded in resin. Sections were cut at 500–700 Å, stained with lead citrate and uranyl acetate, and viewed in an electron microscope. The muscle in *b* was stained for cholinesterase⁹ after aldehyde fixation, to facilitate location of synaptic sites; irregular spots are cholinesterase reaction product.

Third, nerve terminals at normal neuromuscular junctions are capped by processes of Schwann (glial) cells³. In the mutant, in contrast, Schwann cell processes frequently penetrated the synaptic cleft, separating the nerve terminals from BL and muscle fibre. In a few other areas, the BL was attenuated or absent, leaving the nerve terminal and muscle fibre membranes in direct apposition. As a result, whereas nearly all of the nerve terminal membrane is directly apposed to synaptic BL in normal animals, this was true for only 22–35% of the contact area in the mutant (Table 1). Finally, junctional folds indent the post-synaptic membrane in normal mice; these are absent at birth, but appear during the first postnatal week. In contrast, mutant postsynaptic membrane was almost devoid of junctional folds, even in those areas that were apposed to nerve terminals. All of these defects were seen in continuously active respiratory muscles and were as marked at P8 (when the mutants were apparently



healthy) as at P15. Thus, the structural abnormalities in the synapses of s-laminin mutant mice are unlikely to be secondary consequences of poor health or inactivity.

To test the functional consequences of these structural defects, we recorded intracellularly from muscle fibres. Resting potentials were similar in mutants and littermate controls (mean, 54 mV, $n=90$ fibres in mutants; mean, 46 mV, $n=80$ fibres in controls). Miniature endplate potentials (m.e.p.ps; Fig. 3*a, b*), which are believed to reflect spontaneous release of neurotransmitter by individual synaptic vesicles, had similar rise and fall times in mutants and controls (average of 1.3 ms from baseline to peak and 3.6 ms from peak to baseline in mutants; 1.2 and 3.4 ms in controls), suggesting that lack of s-laminin had little effect on the passive membrane properties of the muscle fibres. The amplitude of m.e.p.ps was slightly lower in mutants than in the controls (0.59 ± 0.06 compared with 0.77 ± 0.07 mV; $n=28$ for

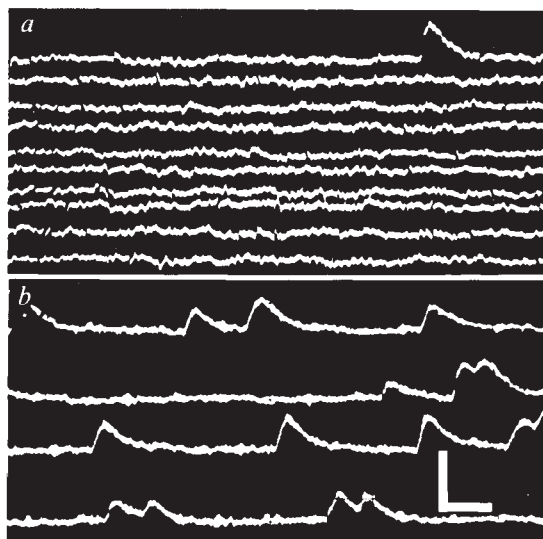
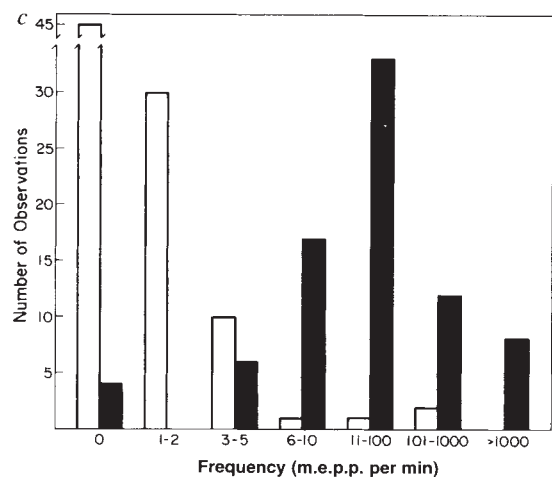


FIG. 3 Spontaneous miniature endplate potentials (m.e.p.ps), recorded intracellularly from muscle fibres in mutant (a) and control (b) muscles. (c) M.e.p.ps are far less frequent in mutants (open bars) than in controls (filled bars). Bars in b show 5 ms and 1 mV.

METHODS. Diaphragms, attached to the rib cage, were dissected from P14–18 mice, and pinned to Sylgard-coated dishes in DMEM medium. The muscle was transilluminated and viewed with a dissecting micro-



scope, and a glass microelectrode was positioned to impale fibres just on the rib's side of the nerve, where almost all neuromuscular junctions are located⁸. The microelectrode was stepped along the muscle in ~0.5-mm intervals between recordings to sample fibres throughout the muscle. Electrodes were filled with 3 M KCl, and had resistances of 40–100 mohms; signals were photographed directly from the oscilloscope screen. Data in c are from four mutant and three control muscles; similar results were obtained but not tallied from an additional three mutants and three controls.

mutant and 15 for control fibres), suggesting either that the density of AChRs in the postsynaptic membrane was slightly affected by s-laminin deficiency, or that subnormal amounts of transmitter reach the postsynaptic membrane.

Most striking, however, was the observation that m.e.p.p. frequency was greatly reduced in mutants (Fig. 3c). No m.e.p.ps at all were observed in half of the mutant muscle fibres during a 1-min observation period (45/90 fibres), even though m.e.p.ps were observed in 95% (76/80) of control fibres. Even excluding

fibres from which no m.e.p.ps were recorded, the frequency was 15-fold lower in mutants than in controls (means of 19 versus 320 per min). This decrease is most simply explained by a reduced capability for transmitter release in the mutant: the dispersal of vesicles could reduce their availability, the paucity of active zones implies a decreased number of release sites, and transmitter released beneath Schwann cell processes is probably unable to reach the postsynaptic membrane. It seems likely that these presynaptic defects also diminish evoked release, contributing to the weakness and eventual death of the mutants.

As a first step in elucidating how s-laminin deficiency leads to synaptic abnormality, we stained unfixed sections of wild-type and mutant muscles with a panel of antibodies against molecules concentrated at the neuromuscular junction (Table 2). Distribution of most antigens was qualitatively normal in the mutant. Although s-laminin was, of course, absent from the synaptic cleft, at least four other components of synaptic BL were present: acetylcholinesterase, collagen $\alpha 5$ (IV), GalNAc-terminated carbohydrate⁵ (assayed with a GalNAc-specific lectin²⁴), and laminin (assayed with polyclonal antiserum that recognizes A, B1 and B2 subunits). Interestingly, the replacement of fetal-type (containing a γ -subunit) AChRs by adult-type (ϵ -subunit-containing) AChRs occurred on schedule¹³ during the first two postnatal weeks, even though the absence of junctional folds had suggested immaturity of the postsynaptic membrane. Thus, several aspects of postsynaptic development occur normally in the mutant. Indeed, we found only one molecular difference between the postsynaptic membrane/cytoskeleton of wild-type and mutant junctions. The neural-cell adhesion molecule N-CAM, initially present throughout the muscle fibre membrane, becomes concentrated at normal endplates during the first two postnatal weeks, but remained at low levels in the mutant. In normal muscles, N-CAM is concentrated at the bases of junctional folds¹⁴; its paucity in the mutant could either reflect or cause the near-absence of folds.

Of presynaptic markers, three intrinsic proteins of synaptic vesicles (SV2, p38/synaptophysin and p65/synaptotagmin) were present at similar levels in mutant and wild-type nerve terminals (Fig. 1k, l), consistent with ultrastructural data presented above. On the other hand, levels of synapsin-like immunoreactivity were

TABLE 2 Immunochemical analysis of endplates in s-laminin mutant mice

Location	Component	Ref	Concentration at endplates			
			P2	P8	P15	P19
Presynaptic membrane	Syntaxin/HPC1	20	+	+	+	+
Synaptic vesicles	SV2	21	+	+	+	+
	Synaptophysin/p38	22	+	+	+	+
	Synaptotagmin/p65	23	ND	+	+	+
	Synapsin (anti-I)	15	±	↓	↓	↓
	(anti-Ia and IIa)		±	↓	↓	↓
Synaptic cleft	(anti-I and II)		±	↓	↓	↓
	S-laminin	4	↓	↓	↓	↓
	GalNAc (VVA-B4)	24	+	+	+	+
	Acetylcholinesterase	9	+	+	+	+
	Collagen $\alpha 5$ (IV)	25	ND	+	+	ND
Postsynaptic membrane	Laminin	4	+	+	+	+
	AChR (α -bungarotoxin)	11	+	+	+	+
	AChR γ -subunit	13	+	+	+	ND
	AChR ϵ -subunit	13	–	+	+	ND
Postsynaptic Cytoskeleton	N-CAM	14	ND	±	↓	↓
	43K/rapsyn	26	+	+	+	+
	59K/syntrophin	27	ND	+	+	+
	DRP/utrophin	26	ND	+	+	+

See Fig. 1 for examples and legend for methods. Antibodies and other stains are described in the references cited, except for polyclonal antisera to synapsin, which were generously provided by P. Greengard and T. Sudhof. See ref. 3 for a general description of the molecular architecture of the synapse. In all cases, littermate controls (wild type and/or heterozygote) were stained in parallel with the mutants. +, Mutant and normal were intensely positive; ±, mutant and normal faintly positive; –, mutant and normal both negative; ↓, normal positive and mutant reduced (N-CAM and synapsin) or absent (s-laminin); ND, not done.

reduced in unfixed mutant nerve terminals (Fig. 1*m, n*). Synapsins are a family of phosphoproteins specifically associated with the cytoplasmic surface of synaptic vesicles¹⁵. Levels of both synapsins (I and II) remained low in mutants in the oldest surviving animals, even though synapsin was readily detectable in normal motor nerve terminals at birth. In contrast, all of the anti-synapsin antibodies stained mutant nerve terminals in fixed sections (Fig. 1*o, p*). Staining was more intense in wild-type than in mutant terminals, but nearly all mutant endplates were distinctly synapsin-positive in sections that had been fixed before they were stained. We suggest that the lack of s-laminin leads to a defect in the structure of synapsin (such as abnormal phosphorylation¹⁵) or in its association with other components of the nerve terminal (such as a cytoskeletal element¹⁶), and that this defect is manifested as an increased extractability of synapsin from unfixed tissue.

In summary, we have presented structural, functional and immunochemical evidence that motor nerve terminals are abnormal in the absence of s-laminin *in vivo*. Together with previous data showing that s-laminin affects neurite outgrowth from motor neurons *in vitro*^{6,7}, these results support the hypothesis that s-laminin is a muscle-derived regulator of nerve-terminal differentiation. The intra-terminal events that s-laminin affects remain to be determined. One possibility is that interactions of s-laminin with the nerve terminal membrane promote assembly of active zones. Lack of active zones in the mutant would almost certainly result in decreased transmitter release, and could also lead to a disruption of cytoskeletal elements that normally interact, directly or via vesicles, with active zones. One such element is synapsin, which binds both to synaptic vesicles and to actin filaments¹⁵, and is normally concentrated near release sites¹⁶. Destabilization of synapsin could lead in turn to dispersal of vesicles. This idea is consistent with results from synapsin-I-deficient mutant mice: paired pulse facilitation of transmitter release was enhanced in this mutant, suggesting that synapsin normally anchors vesicles at or near release sites¹⁷. An alternative mechanism is suggested by the hypothesis that synapsins regulate formation as well as function of synapses¹⁵. Thus, s-laminin could act by promoting the association of synapsins with other intra-terminal components.

Unexpectedly, we found that not only the nerve terminal, but also the muscle fibre and Schwann cell, were affected in the mutant: the postsynaptic apparatus has little N-CAM and few folds, and long Schwann cell processes extend into the synaptic cleft. The presence of abnormalities in all three cells of the neuromuscular junction raises the question of which cells are direct targets of s-laminin. For example, if s-laminin promotes nerve-terminal differentiation^{6,7}, then Schwann cells might recognize mutant nerve terminals as abnormal and envelop them, as in some diseased, aged or damaged muscles^{18,19}. The undifferentiated and unwrapped terminals might then be unable to provide the muscle with signals that trigger generation of folds. Alternatively, the signalling pathway might lead from BL to either muscle or Schwann cell and then to nerve, or even from BL separately to all three cells. It is also possible that s-laminin anchors other molecules in the BL, and that their loss contributes to the defects we have documented. In any event, our results clearly demonstrate that s-laminin is important in synaptic differentiation, and provide a basis for testing these and other alternatives. More generally, the observation that discrete aspects of synaptic maturation are selectively perturbed in s-laminin-deficient mice suggests that a molecular genetic analysis of synaptogenesis will be feasible in mammals. □

Received 31 October 1994; accepted 24 January 1995.

- Sanes, J. R., Marshall, L. M. & McMahan, U. J. *J. Cell Biol.* **78**, 176–198 (1978).
- Burden, S. J., Sargent, P. B. & McMahan, U. J. *J. Cell Biol.* **82**, 412–425 (1979).
- Hall, Z. W. & Sanes, J. R. *Cell* **72**/Neuron **10** (suppl.) 99–121 (1993).
- Sanes, J. R., Engvall, E., Butkowski, R. & Hunter, D. D. *J. Cell Biol.* **111**, 1685–1699 (1990).
- Green, T. L., Hunter, D. D., Chan, W., Merlie, J. P. & Sanes, J. R. *J. biol. Chem.* **267**, 2014–2022 (1992).

- Hunter, D. D. *et al. Cell* **59**, 905–913 (1989).
- Porter, B. E., Weis, J. & Sanes, J. R. *Neuron* (in the press).
- Merlie, J. P. & Sanes, J. R. *Nature* **317**, 66–68 (1985).
- Karnovsky, M. J. *J. Cell Biol.* **23**, 217–232 (1964).
- Gurney, M. E., Yamamoto, H. & Kwon, Y. J. *Neurosci.* **12**, 3241–3247 (1992).
- Balice-Gordon, R. J. & Lichtman, J. W. *J. Neurosci.* **13**, 834–855 (1993).
- Balice-Gordon, R. J., Chua, C. K., Nelson, C. C. & Lichtman, J. W. *Neuron* **11**, 801–815 (1993).
- Gu, Y. & Hall, Z. W. *Neuron* **1**, 117–125 (1988).
- Covault, J. & Sanes, J. R. *J. Cell Biol.* **102**, 716–730 (1986).
- Greengard, P., Valtorta, F., Czernik, A. J. & Benfenati, F. *Science* **259**, 708–784 (1993).
- Hirokawa, N., Sobue, K., Kanda, K., Harada, A. & Yonifuji, H. *J. Cell Biol.* **108**, 111–126 (1989).
- Rosahl, T. W. *et al. Cell* **75**, 661–670 (1993).
- Hutchinson, D. O. *et al. Brain* **116**, 633–653 (1993).
- Jirmanová, I. *J. Neurocyt.* **4**, 141–155 (1975).
- Inoue, A., Obata, K. & Akagawa, K. *J. biol. Chem.* **267**, 10613–10619 (1992).
- Buckley, K. & Kelly, R. B. *J. Cell Biol.* **100**, 1284–1294 (1985).
- Südhof, T. C., Lottspeich, F., Greengard, P., Ehrenfried, M. & Jahn, R. *Science* **238**, 1142–1144 (1987).
- Matthew, W. D., Tsavaler, L. & Reichardt, L. F. *J. Cell Biol.* **91**, 257–269 (1981).
- Scott, L. J. C., Bacou, F. & Sanes, J. R. *J. Neurosci.* **8**, 932–944 (1988).
- Miner, J. H. & Sanes, J. R. *J. Cell Biol.* **127**, 879–891 (1994).
- Phillips, W. D., Maimone, M. M. & Merlie, J. P. *J. Cell Biol.* **115**, 1713–1723 (1991).
- Ervasti, J. M. & Campbell, K. P. *Cell* **66**, 1121–1131 (1991).
- McMahan, A. P. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
- Nagy, A., Rossant, J., Nagy, T., Abramow-Newerly, W. & Roder, J. C. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8424–8428 (1993).
- Hunter, D. D., Shah, V., Merlie, J. P. & Sanes, J. R. *Nature* **338**, 229–234 (1989).

ACKNOWLEDGEMENTS. We thank M. Nichol, B. Klocke and J. Cunningham for assistance; A. Nagy for R1-ES cells; our colleagues for gifts of antisera (see Table 2); and A. Missias for affinity-purifying anti-AChR antibodies. This work was supported by grants from the NIH and the Muscular Dystrophy Association (USA) to J.P.M. and J.R.S., and the Council for Tobacco Research to J.P.M. P.G.N. was supported by fellowships from the Medical Research Council (Australia) and the Muscular Dystrophy Association (USA). P.G.N. and M.G. contributed equally to this work.

Nitric oxide mediates activity-dependent synaptic suppression at developing neuromuscular synapses

Ti Wang, Zuoping Xie* & Bai Lu†

Roche Institute of Molecular Biology, Nutley, New Jersey 07110, USA

TEMPORAL correlation between pre- and postsynaptic activities is an important mechanism that regulates synaptic connectivity during development and synaptic plasticity in the adult^{1–3}. In developing neuromuscular junctions, postsynaptic activity is critical in functional suppression and, ultimately, elimination of the synapses^{4–6}. Although repetitive postsynaptic firing asynchronous to the presynaptic activity results in a persistent synaptic suppression^{7–10}, the underlying molecular mechanism remains unknown. Here we provide evidence that nitric oxide (NO), a free radical implicated in several forms of synaptic plasticity^{11–15}, may serve as a retrograde signal for activity-dependent suppression in the neuromuscular synapse. NO donors and activators of the cyclic GMP pathway suppressed spontaneous and evoked synaptic currents. Moreover, the synaptic suppression induced by repetitive postsynaptic depolarization was prevented by the NO-binding protein haemoglobin and by inhibitors of NO synthase. Thus, synaptic suppression may be triggered by NO released from a postsynaptic myocyte that fires asynchronously to the presynaptic terminal.

Synaptic currents were monitored on single innervated myocytes by whole-cell voltage-clamp recordings in 1-day-old nerve-muscle cultures prepared from *Xenopus* embryos¹⁶. We first examined the effects of the NO donor, *S*-nitroso-*N*-acetylpenicillamine (SNAP)¹⁷, on spontaneous synaptic currents (s.s.cs). Bath application of SNAP elicited a 52% decrease in

* On leave from Tsinghua University, Beijing, China.

† To whom correspondence should be addressed.

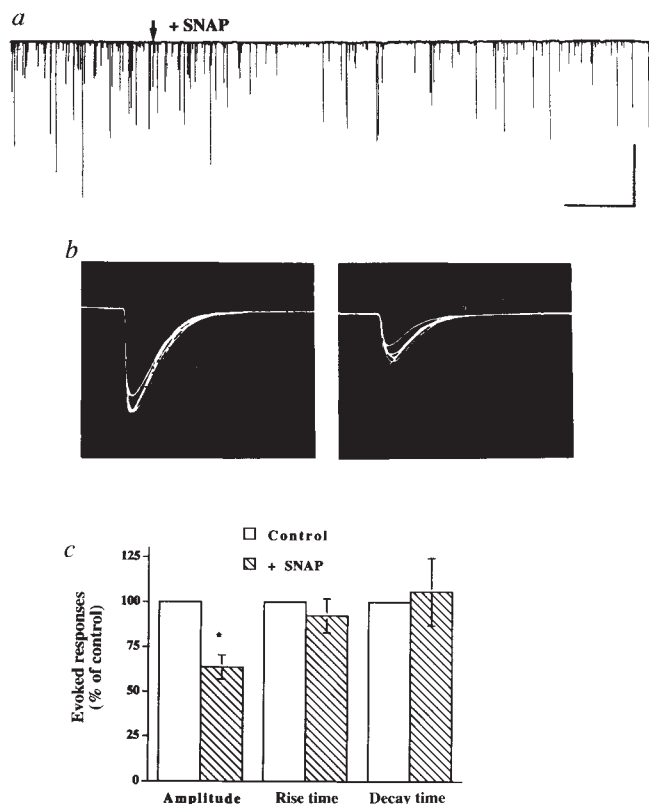


FIG. 1 Effects of the NO donor SNAP (100 μ M) on spontaneous and evoked synaptic currents (s.s.c.s and e.s.c.s) in 1-day-old nerve-muscle cultures. *a*, Continuous trace depicting s.s.c.s recorded from an innervated myocyte before and after bath application of SNAP. Downward deflections are s.s.c.s ($V_h = -70$ mV, filtered at 150 Hz). Note the changes in s.s.c. frequency. *b*, Oscilloscope traces depicting five consecutively superimposed e.s.c.s before (left) and 8 min after (right) SNAP treatment (filtered at 3 kHz). Note the decrease in e.s.c. amplitudes after SNAP treatment. Scales, 0.5 nA, 4 min (*a*) and 1.2 nA, 28 ms (*b*). *c*, Effects of SNAP on the properties of e.s.c.s. Rise time refers to interval between 10% and 90% of peak e.s.c. amplitude. Decay phase of e.s.c.s was fitted by a single exponential curve and decay time is defined as the time need for e.s.c. amplitude to drop to 1/e of the peak value. The averaged values after SNAP treatment were normalized to control values for each synapse, and data were then pooled and calculated for mean $\pm$ s.e.m. *, $P < 0.05$, two-tailed *t*-test; $n = 6$.

METHODS. The neural tube and the associated myotomal tissue of stage 20–22 *Xenopus* embryos were dissociated in Ca^{2+} - and Mg^{2+} -free saline supplemented with EDTA. The cells were plated on glass coverslips and grown at room temperature for 1 day before use. The culture medium consisted of 50% (vol/vol) Ringer's solution (115 mM NaCl, 2 mM CaCl_2 , 2.5 mM KCl, 10 mM HEPES, pH 7.6), 49% L-15 Leibovitz medium (Sigma), and 1% fetal bovine serum (GIBCO). Synaptic currents (both s.s.c.s and e.s.c.s) were recorded from innervated myocytes at room temperature in culture medium under the whole-cell configuration, with voltage clamped at -70 mV. The pipette solution contained 150 mM KCl, 1 mM NaCl, 1 mM MgCl_2 , and 10 mM HEPES (pH 7.2). To elicit e.s.c.s, presynaptic neuronal soma was stimulated to fire action potentials by a heat-polished patch electrode containing Ringer's solution. All data were collected by a patch-clamp amplifier (EPC-7), with a current signal filter at 3 kHz. The data were stored on a videotape recorder for later playback on a storage oscilloscope (Textonic TDS 420) and a chart recorder (Gould EasyGraf 240). The amplitude, rise and decay times of s.s.c.s were analysed by SCAN program (Dagan Inc.).

s.s.c. frequency (Figs 1*a* and 2*e*). The effect of SNAP was dose-dependent (effector concentration for half-maximum response, $\text{EC}_{50} \sim 50$ μ M), reversible within 20 min (see below) and relatively fast (Fig. 1*a*). To determine whether NO also regulates functional synaptic transmission, we examined the effects of the

NO donor on impulse-evoked synaptic currents (e.s.c.s). Treatment with SNAP resulted in a 37% decrease in the amplitudes of e.s.c.s, whereas there was no change in e.s.c. rise and decay times (Fig. 1*b*, *c*).

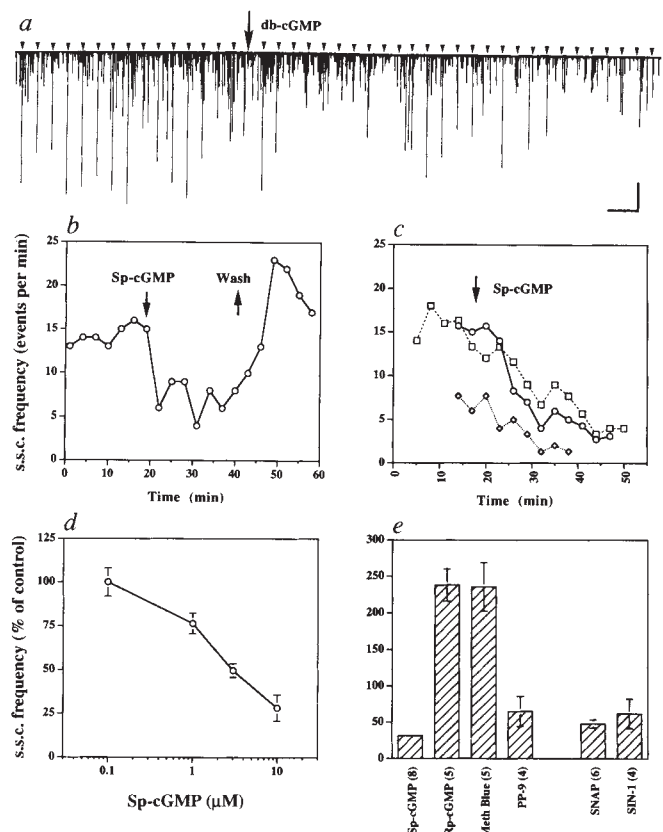
The NO effects could be mediated by cGMP-dependent or cGMP-independent pathways, through interaction of NO with metal- or thiol-containing proteins^{18,19}. NO has been shown to collapse neuronal growth cones through a non-cGMP mechanism²⁰. To examine the mechanism underlying NO-induced synaptic suppression, we treated the neuromuscular synapses with the membrane-permeable, specific cGMP-dependent protein kinase (G-kinase) activator, dibutyl cGMP (db-cGMP). Bath application of the cGMP analogue significantly decreased the e.s.c. amplitude and s.s.c. frequency (Fig. 2*a*; $\text{EC}_{50} \sim 0.1$ mM, $n = 4$). Sp-8-pCPT-cGMPS (Sp-cGMP), a more potent cGMP analogue, elicited a 68% decrease in s.s.c. frequency (Fig. 2*e*). Action of Sp-cGMP was also rapid, reversible and dose-dependent (Fig. 2*b–d*).

Several reagents known to interfere with various aspects of the NO/cGMP pathway were examined. Like SNAP, the NO donors 3-morpholiniosydnonimine HCl (SIN-1) (Fig. 2*e*) and sodium nitroprusside (not shown) decreased both e.s.c. amplitude and s.s.c. frequency. Rp-8-pCPT-cGMPS (Rp-cGMP), a G-kinase inhibitor, increased the s.s.c. frequency (Fig. 2*e*) and the e.s.c. amplitude (not shown). The formation of cGMP is catalysed by cytosolic guanylate cyclase. Protoporphyrin-9 (PP-9), a guanylate cyclase activator²¹, reduced s.s.c. frequency by 40%, whereas the enzyme inhibitor methylene blue²² had the opposite effect (Fig. 2*e*). Moreover, the effects of SNAP and Sp-cGMP occluded each other ($n = 4$), and inhibition of G-kinase by Rp-cGMP blocked the suppression induced by SNAP ($n = 6$). Although we cannot rule out the possibility that a cGMP-independent mechanism might contribute to the decrease of synaptic activity owing to NO-triggered collapse of nerve terminals²⁰, our data support the idea that the NO effects are mediated by activation of guanylate cyclase, and subsequently cGMP-dependent protein phosphorylation.

The decrease in s.s.c. frequency could result either from fewer acetylcholine (ACh) vesicles being released presynaptically, or from a decrease in postsynaptic ACh sensitivity. The latter usually parallels a general reduction of s.s.c. amplitude, and/or changes in s.s.c. rise and decay times²³. No significant change was found in any of these parameters after SNAP treatment (*t*-test, $P > 0.1$). Moreover, SNAP did not alter the ACh sensitivity on the myocyte surface (measured by the amplitudes of ACh iontophoretic currents, $n = 6$), nor did it change the decay time of e.s.c.s (Fig. 1*c*). These data argue for a presynaptic effect of NO. However, Sp-cGMP treatment resulted in a 32% decrease in the average amplitude, and a 21% increase in rise time of s.s.c.s (*t*-test, $P < 0.05$). Thus, Sp-cGMP may cause a postsynaptic reduction of ACh responsiveness in addition to the presynaptic effects.

Repetitive postsynaptic depolarization, which leads to muscle action potentials asynchronous to the presynaptic activity, can produce synaptic suppression (as reflected by a decrease of s.s.c. frequency and e.s.c. amplitude)¹⁰. The induction of the suppression requires postsynaptic Ca^{2+} influx¹⁰, which activates NO synthase¹⁸. If endogenous NO from muscle cells is a retrograde signal that initiates the suppression, it should be abolished by inhibiting NO synthase or by absorbing NO with haemoglobin. We induced the synaptic suppression by repetitive, depolarizing stimulation of the postsynaptic myocyte with the recording pipette under current-clamp configuration (8 Hz, 15 min) (Fig. 3*a*, upper trace). The NO synthase inhibitor N^G-mono-methyl-L-Arginine (NMMA)²⁴ was then added to the culture dish, and the same experiment was done on another innervated myocyte. NMMA (100 μ M) virtually abolished the synaptic suppression induced by postsynaptic depolarization (Fig. 3*a*, lower trace). Statistical analysis indicated that postsynaptic depolarization decreased the mean e.s.c. amplitude by 55% in control synapses,

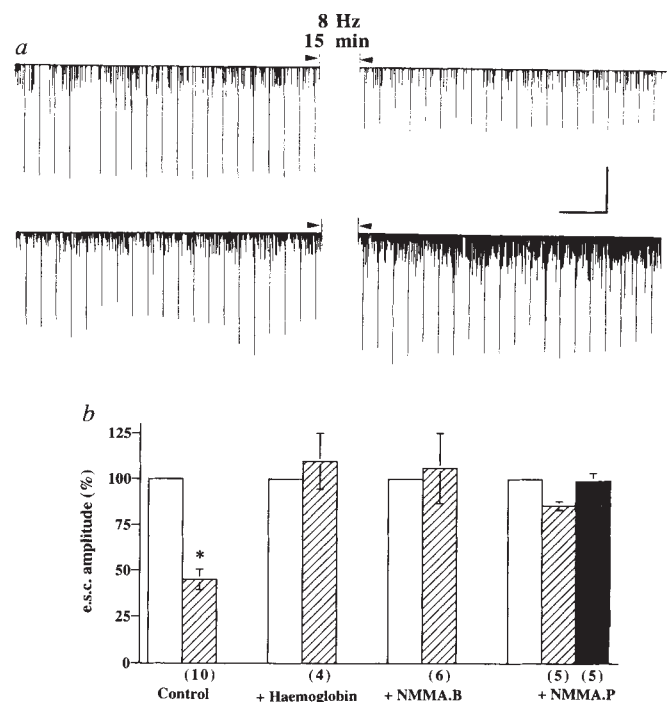
FIG. 2 Involvement of the cGMP pathway and NO in synaptic suppression. *a*, Inhibition of s.s.c.s and e.s.c.s by db-cGMP. E.s.c.s appear as large, regularly spaced downward deflections (indicated by arrowheads), amid small and randomly occurring s.s.c.s. Note that after bath application of db-cGMP (1 mM), both the amplitude of e.s.c.s and the frequency of s.s.c.s are reduced. *b*, Reversibility of Sp-cGMP effects on s.s.c. frequency. Arrows indicate perfusion of Sp-cGMP (final concentration: 10 μ M) into and out of a culture dish. *c*, Time course of Sp-cGMP effects. Recordings of 3 individual synapses are shown. Each point was calculated as s.s.c. frequency averaged from 3 min of recording in both *b* and *c*. *d*, Concentration dependence of Sp-cGMP effects. Sp-cGMP was applied repeatedly to a culture to achieve the given concentrations. S.s.c. frequencies (averaged from 4 min of recording) were measured 7–10 min after application of Sp-cGMP at each concentration, and normalized to that before drug treatment. Error bars indicate standard errors. *e*, Comparison of drugs that affect various aspects of the NO/cGMP pathway. Drugs were added to the culture media after a period of control recording. Spontaneous synaptic current frequencies were measured in pairs for the same synapse over equal duration before and 10 min after drug treatments. The ratios of s.s.c. frequencies before and after drug treatments were calculated for each individual synapse. All values represent mean $\pm$ s.e.m. Final concentrations of the drugs are: Sp-cGMP, 10 μ M; Rp-cGMP, 10 μ M; methylene blue (Meth. Blue), 20 μ M; Protoporphyrin-9 (PP-9), 10 μ M; SNAP, 100 μ M; 3-morpholinodimethylamine HCl (SIN-1), 250 μ M. Total number of experiments for each drug is indicated under each column. All changes (increase or decrease after drug treatments) are statistically significant compared with control ($P < 0.05$, two-tailed *t*-test).



but had no effect on synapses treated with NMMA (Fig. 3*b*). Similar experiments were done using freshly prepared haemoglobin (30 μ M)²⁵. The mean e.s.c. amplitude after depolarization was 45% in the control and 109% in the haemoglobin-treated, paired synapses (Fig. 3*b*). Haemoglobin and NMMA did not affect basal s.s.c. frequency (not shown). To determine the loca-

tion of the NO synthase activity, we delivered NMMA directly into postsynaptic myocytes through the whole-cell recording pipette. With 100 μ M NMMA in the pipette, synaptic suppression was virtually reversed (Fig. 3*b*, hatched bar). A higher NMMA concentration in the pipette (300 μ M) completely prevented the activity-dependent suppression (Fig. 3*b*, black bar). These data,

FIG. 3 Effects of haemoglobin and the NO synthase inhibitor NMMA on synaptic suppression induced by postsynaptic depolarization. *a*, Examples of evoked synaptic currents (e.s.c.s) recorded from two synapses treated with or without NMMA. Upper trace, presynaptic neuron was stimulated (0.05 Hz), and synaptic efficacy was measured by the amplitudes of e.s.c.s recorded from an innervated myocyte ($V_h = -70$ mV, filtered at 150 Hz). The e.s.c.s appear as large, regularly spaced downward deflections amid randomly occurring small s.s.c.s. Action potentials were then induced in the postsynaptic myocyte by repetitive depolarizing pulses (8 Hz, 15 min) delivered by the same patch pipette switched to current-clamp mode. Note the dramatic reduction of the e.s.c. amplitudes after repetitive depolarization. Lower trace, NMMA (100 μ M) was added to the culture and the same experiment was done on another innervated myocyte. In the presence of NMMA, there was virtually no change in the amplitudes of e.s.c.s after repetitive depolarization. Scales, 3 nA, 1 min (upper trace); 1.5 nA, 1 min (lower trace). *b*, Summary of the effects of haemoglobin and NMMA. The mean e.s.c. amplitudes before and after depolarization were calculated, and the values were normalized to those before depolarization for each synapse. The data were then averaged and presented as percentages of the values before depolarization. Error bars indicate standard errors. *, $P < 0.001$, two-tailed *t*-test. The open bars are average e.s.c. amplitudes before depolarization and hatched and black bars are those after depolarization. The numbers associated with the data represent the total number of experiments performed. Haemoglobin, 30 μ M. NMMA.B, NMMA added to culture media at a final concentration of 100 μ M. NMMA.P, NMMA applied directly to postsynaptic myocytes with patch pipettes at 100 μ M (hatched bar) or 300 μ M (black bar). **METHODS.** Commercial haemoglobin contains a substantial amount of the oxidized derivative methaemoglobin, and therefore its binding capacity to NO is low. Pure haemoglobin (oxyhaemoglobin) was prepared by treatment with the reducing agent $\text{Na}_2\text{S}_2\text{O}_4$ (10-fold molar excess), which was then removed by dialysis against Ringer's solution. The solution was frozen in aliquots and stored at -80°C for up to 2 weeks²⁵.



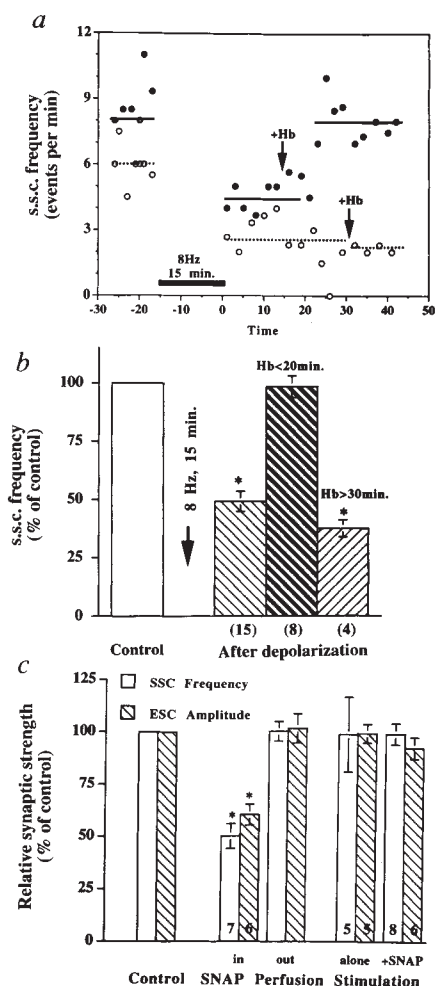


FIG. 4 *a*, Examples of haemoglobin effects on s.s.c. frequency after the induction of synaptic suppression. Each data point was calculated as s.s.c. frequency averaged from 2–3 min of recording. Synaptic suppression was induced by the same methods described in Fig. 3. Horizontal lines indicate mean s.s.c. frequencies. Note that the depression of s.s.c. frequency was almost completely recovered when haemoglobin (Hb) was added to the culture 12 min after termination of postsynaptic stimulation (filled circle), but Hb was virtually ineffective when added 32 min after the stimulation (open circle). *b*, Summary of the haemoglobin effects. Haemoglobin was added to the culture either less than 20 min (Hb < 20 min) or more than 30 min (Hb > 30 min) after the termination of postsynaptic stimulation. The ratios of the mean s.s.c. frequencies before and after synaptic suppression (no treatment, Hb < 20 min and Hb > 30 min) were calculated for each synapse. The data were then averaged and presented as percentages of control (mean $\pm$ s.e.m.). Total numbers of experiments are indicated under each bar. *, statistically different from control at $P < 0.001$, two tailed t-test. *c*, Effects of NO on active nerve terminals. A rapid and reversible inhibition of s.s.c. frequency was achieved by local perfusion with a SNAP-containing pipette (2 mM) for 1–2 min ('in' and 'out', under appropriate bars). The presynaptic neurons were stimulated (2–8 Hz, 2 min) either alone or with brief application of SNAP ('alone' or '+SNAP'). The ratios of s.s.c. frequency (open bars) or e.s.c. amplitude (hatched bars) before and after presynaptic stimulation and/or SNAP perfusion were calculated for each individual synapse, and averaged values are represented as percentages of control (before stimulation and/or SNAP treatment, mean $\pm$ s.e.m.). The numbers associated with the data represent the total number of experiments performed. *, $P < 0.001$, two tailed t-test.

together with the observation that NO donors inhibited synaptic activity independent of postsynaptic depolarization and/or Ca^{2+} influx, strongly suggest that endogenous NO is a signal that triggers synaptic suppression, and that the signal is generated from postsynaptic myocytes by repetitive depolarization.

The synaptic suppression appears to be long lasting^{9,10}. A detailed analysis using a fast perfusion mechanism indicated that although prolonged treatment of SNAP (>30 min) resulted in a persistent suppression of s.s.c. frequency ($n=4$), short-term NO effects (within 20 min) were reversible. Moreover, the decrease in s.s.c. frequency induced by postsynaptic stimulation could be reversed as long as the haemoglobin was applied to the culture within 20 min after termination of the stimulation (Fig. 4*a, b*). Similar effects were observed for e.s.c. amplitude (not shown). The NO scavenger became ineffective when applied 30 min after the stimulation. Thus, NO may serve as an initial, reversible retrograde messenger that triggers the downstream events (such as structural changes and regulation of gene expression) responsible for the maintenance of the persistent suppression.

Active nerve terminals are protected from synaptic suppression induced either by tetanic stimulation of one of the two coinnervating, presynaptic motor neurons, or by ACh pulses directly onto the myocyte surface^{7,8}. More active terminals are less susceptible to the depression induced by prolonged stimulation of the postsynaptic myocytes¹⁰. However, because the developing spinal neurons are very delicate and cannot tolerate the prolonged stimulation, it was technically difficult to show that synchronized pre- and postsynaptic stimulation prevents synaptic suppression. To determine whether the active terminals are also spared from the depressive effects of NO, we took advantage of rapid and short-term reversible actions of the NO donor

SNAP. A very brief exposure to SNAP (by perfusion with a SNAP-containing pipette for 1–2 min) caused a transient and reversible decrease of s.s.c. frequency and e.s.c. amplitude (Fig. 4*c*). The depressive effects of SNAP can be prevented if the cultures are pretreated with haemoglobin. Simultaneous application of SNAP and presynaptic stimulation prevented the decrease of s.s.c. frequency and e.s.c. amplitude (Fig. 4*c*). Stimulation of the presynaptic neurons alone did not affect the synaptic efficacy. Thus, active nerve terminals appear to be protected from the NO effects.

Our findings, together with previous studies of NO function in long-term potentiation and depression (LTP and LTD) of synaptic efficacy in the adult brain^{11–13,15,26}, suggest a general role of NO as a retrograde signal in the hebbian-type mechanism, including regulation of neuromuscular synaptogenesis during development and brain synaptic plasticity in the adult. LTP is achieved by coactivation of pre- and postsynaptic neurons in the adult hippocampus²⁷. The developing neuromuscular synapses may be suppressed by asynchronous activity between the motor neuron and the muscle cell^{7,10}. Asynchrony of pre- and postsynaptic activity is unavoidable when a myocyte receives multiple innervation during development. An active terminal may trigger synchronous firing in the postsynaptic myocyte, which could release NO and suppress the coinnervating, inactive terminals. In addition to the implication that the muscle-derived NO may serve as a general signal for feedback inhibition, our results may also have relevance to the neuromuscular dissociation seen in extreme cases of muscle fatigue²⁸. □

Received 3 October 1994; accepted 27 January 1995.

1. Stent, G. *Proc. natn. Acad. Sci. U.S.A.* **70**, 997–1001 (1973).
2. Shatz, C. J. *Neuron* **5**, 745–756 (1990).

3. Constantine-Paton, M., Cline, H. T. & Debski, E. A. *Rev. Neurosci.* **13**, 129–154 (1990).
4. Betz, W. J. in *The Vertebrate Neuromuscular Junction* (ed. Salpeter, M. M.) (Liss, New York, 1987).
5. Srihari, T. & Vrbova, G. *J. Neurocytol.* **7**, 529–540 (1978).
6. Ridge, R. M. A. P. & Betz, W. J. *J. Neurosci.* **4**, 2614–2620 (1984).
7. Lo, Y. & Poo, M.-m. *Science* **254**, 1019–1023 (1991).
8. Dan, Y. & Poo, M.-m. *Science* **256**, 1570–1573 (1992).
9. Lo, Y. J. & Poo, M.-m. *J. Neurosci.* **14**, 4684–4693 (1994).
10. Lo, Y. J., Lin, Y. C., Sanes, D. H. & Poo, M.-m. *J. Neurosci.* **14**, 4694–4704 (1994).
11. Bohme, G. A., Bon, C., Stutzmann, J., Doble, A. & Blachard, J. *Eur. J. Pharmac.* **199**, 379–381 (1991).
12. Schuman, E. M. & Madison, D. V. *Science* **254**, 1503–1506 (1991).
13. O'Dell, T. J., Hawkins, R., Kandel, E. R. & Arancio, O. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11285–11289 (1991).
14. Schuman, E. M. & Madison, D. V. *Science* **263**, 532–536 (1994).
15. Shibuki, K. & Okada, D. *Nature* **349**, 326–328 (1991).
16. Lu, B., Greengard, P. & Poo, M. M. *Neuron* **8**, 521–529 (1992).
17. Southam, E. & Garthwaite, J. *Neurosci. Lett.* **130**, 107–111 (1991).
18. Bret, D. S. & Snyder, S. H. *Neuron* **8**, 3–11 (1992).
19. Stamer, J. S. *Cell* **78**, 931–936 (1994).
20. Hess, D. T., Patterson, S. I., Smith, D. S. & Skene, J. H. P. *Nature* **366**, 562–565 (1993).
21. Wolin, M. S. *J. biol. Chem.* **257**, 13312–13320 (1982).
22. Gruetter, C. A., Gruetter, D. Y., Lyon, J. E., Kadowitz, P. J. & Ignarro, L. J. *J. Pharmac. exp. Ther.* **219**, 181–186 (1981).
23. Kidokoro, Y. *Neurosci. Res.* **1**, 157–170 (1984).
24. Izumi, Y., Clifford, D. B. & Zorumski, C. F. *Science* **257**, 1273–1276 (1992).
25. Martin, W., Villani, G. M., Jothianandan, D. & Furchgott, R. F. *J. Pharmac. exp. Ther.* **232**, 708–716 (1985).
26. Crepel, F. & Jaillard, D. *Neuroreport* **1**, 133–136 (1990).
27. Bliss, T. V. P. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
28. Ghez, C. in *Principle of Neural Science* (eds Kandel, E. R., Schwartz, J. H. & Jessell, T. M.) (Appleton & Lange, Norwalk, CT, 1991).

ACKNOWLEDGEMENTS. We thank M.-m. Poo for his valuable advice and constant support throughout this work, J. A. Connor, T. Curran and F. L. Margolis for helpful discussions and critical comments of the manuscript and Y.-J. Lo for advice on some of the experiments.

Control of lamprey locomotor neurons by colocalized monoamine transmitters

Judith Schotland^{*†}, Oleg Shupliakov^{*},
Martin Wikström^{*}, Lennart Brodin^{*},
Meera Srinivasan^{*}, Zhi-bing You[†],
Mario Herrera-Marschitz[†], Wei-qing Zhang^{*},
Tomas Hökfelt^{*} & Sten Grillner^{*§}

The Nobel Institute for Neurophysiology, ^{*} Department of Neuroscience, and [†] Department of Pharmacology, Karolinska Institutet, S-171 77 Stockholm, Sweden

NEURONS in the central nervous system (CNS) often store more than one neurotransmitter^{1,2}, but as yet the functional significance of this type of coexistence is poorly understood. 5-Hydroxytryptamine (5-HT) modulates calcium-dependent K⁺ channels (K_{Ca}) responsible for the postspike afterhyperpolarization in different regions of the CNS^{3,4}. In lamprey, 5-HT neurons control apamine-sensitive K_{Ca} channels in spinal locomotor network interneurons^{4–6}, thereby in addition regulating the duration of locomotor bursts^{7,8}. We report here that these spinal 5-HT neurons also contain dopamine. Like 5-HT, dopamine causes a reduction of the afterhyperpolarization, but in this case it is due to a reduction of calcium entry during the action potential, which results in a reduced activation of K_{Ca}. 5-HT and dopamine are both released from these midline neurons, and both reduce the afterhyperpolarization through two distinctly different, but complementary cellular mechanisms. The net effect of dopamine (10–100 μM) on the locomotor network is similar to that of 5-HT, and the effects of dopamine and 5-HT are additive at the network level.

Figure 1 shows transverse sections of the lamprey spinal cord. In Fig. 1a, b a single section has been double-labelled with

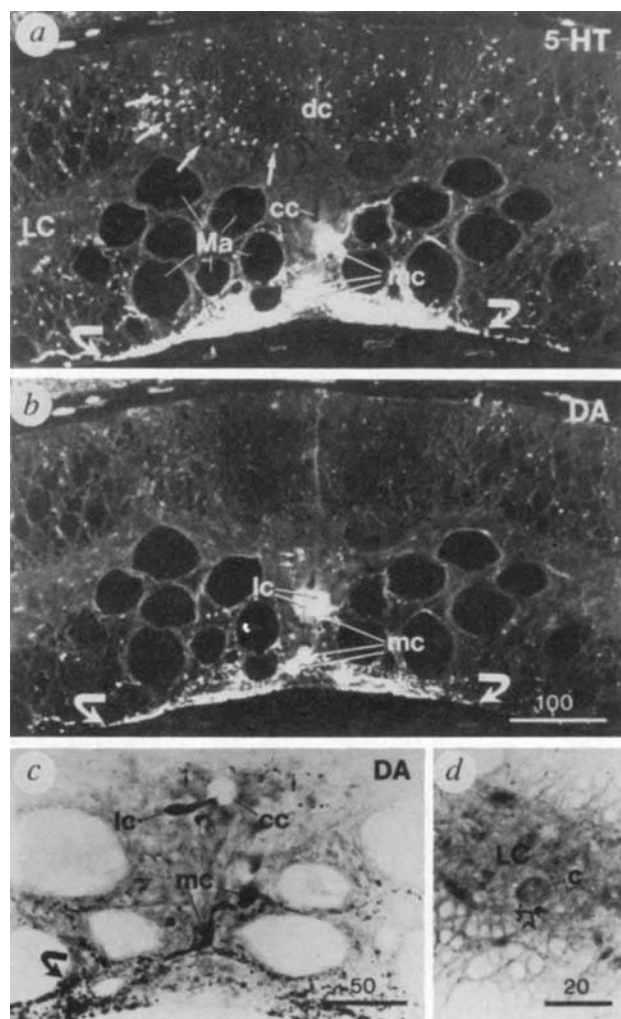


FIG. 1 Dopamine and 5-HT are colocalized in neurons of the ventral plexus in the lamprey spinal cord. a, 5-HT-immunoreactivity is present in three ventromedial cells (mc) below the central canal (cc), as well as in the dense plexus on the ventromedial aspect of the spinal cord below the giant Müller axons (Ma; area between the curved arrows; transverse sections). b, Dopamine-immunoreactivity is present in the same three ventromedial cells (mc) that exhibit 5-HT-immunoreactivity, as well as in fibres in the ventromedial plexus. Note the selective labelling of 5-HT-immunoreactive fibres in the dorsal column (dc, arrows in a), and of two DA-immunoreactive liquor-contacting cells (lc in b) close to the central canal. c, Detailed morphology of DA-immunoreactive cells (PAP-labelling) in the ventromedial spinal cord. Two ventromedial cells (mc) below the central canal (cc) and the ventromedial plexus (curved arrow) show DA immunoreactivity. Labelled liquor-contacting cells (lc) are located immediately ventral to the central canal. Additional labelled processes are clustered immediately below the dorsal columns (small arrows). Scale bar units, μm. d, In the lateral cell column (LC) DA-immunoreactive bouton-like elements occur in the vicinity of unlabelled cell bodies (c). Specificity tests (18) have shown that the DA antiserum does not crossreact with conjugated 5-HT.

METHODS. Spinal cords were fixed in 4% p-formaldehyde + 0.5% glutaraldehyde⁹. Cryostat sections of 14 μm thickness of *Lampetra fluviatilis* (n = 2), and *Ichthyomyzon unicuspis* (n = 1), were incubated in rabbit anti-DA (1:400; ref. 18) and guinea pig anti-5-HT (1:400; ref. 19) antisera, and subsequently in Texas red-conjugated donkey anti-rabbit (1:10; Amersham) and fluorescein isothiocyanate-conjugated goat anti-guinea pig (1:40; Jackson Immunolabs) antibodies. The 5-HT immunoreactivity was strongly reduced in tissue fixed without p-formaldehyde, and conversely the DA immunoreactivity was reduced if glutaraldehyde was omitted. For PAP-immunohistochemistry 3% glutaraldehyde was used. Sections from adult *Lampetra fluviatilis* (n = 3) were incubated in anti-dopamine antiserum (1:1,000–1:2,000), followed by a standard peroxidase anti-peroxidase (PAP) protocol¹⁰.

[†] Present address: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

[§] To whom correspondence should be addressed.

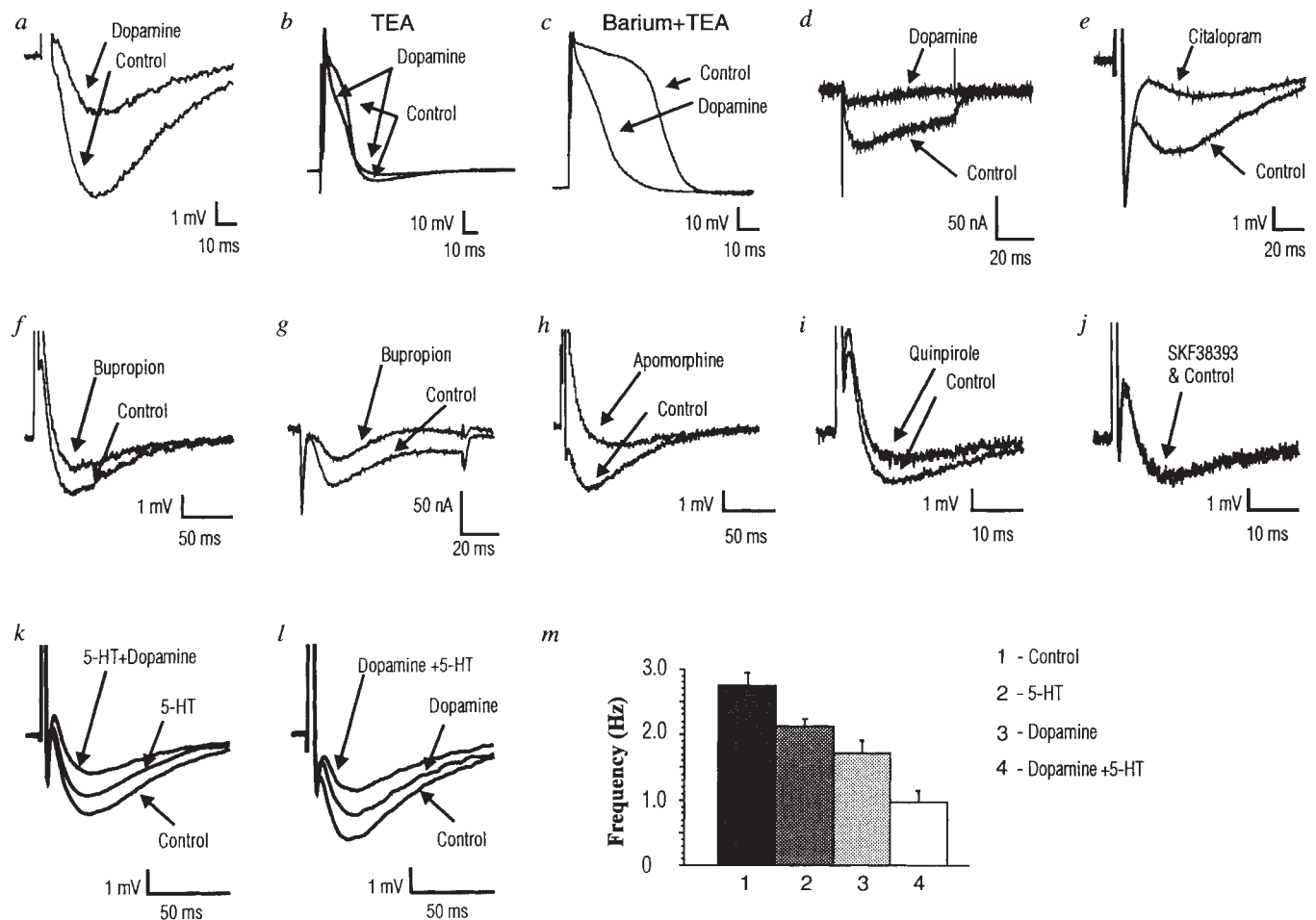


FIG. 2 Dopamine reduces the amplitude of the slow AHP in spinal neurons (a). In the presence of the K^+ channel blocker tetraethylammonium (TEA), DA shortens the action potential duration (b) and reduces the amplitude of the AHP. In the presence of Ba^{2+} (c) the duration of broadened action potentials is markedly reduced by DA. The membrane potential remained unchanged after DA application, and there was no change in membrane conductance as judged from applying hyperpolarizing current pulses (not illustrated). In voltage clamp experiments (d) DA causes a reduction of the Ca^{2+} current. In these experiments, done in the presence of 1.5 μ M tetrodotoxin, 10 mM TEA and 2–3 mM CsCl, the voltage steps were adjusted to evoke a maximal inward current response. These currents were identified as Ca^{2+} currents by application of 400 μ M $CdCl_2$. Bath-application of the selective 5-HT re-uptake blocker citalopram (2 μ M) (e) caused a depression of the AHP in 4/4 cells tested. Another 5-HT re-uptake blocker, zimelidine, has previously been shown to elicit this effect in a proportion of the cells tested²⁰. The DA re-uptake blocker, bupropion (100–200 μ M), also reduces the AHP under current-clamp conditions (f). Under voltage-clamp (g), bupropion causes a corresponding reduction of the Ca^{2+} current, further corroborating that endogenously released DA acts through a suppression of the Ca^{2+} current. The D1/D2 receptor agonist, apomorphine (h), and the D2/D3 receptor agonist, quinpirole (i), reduces the amplitude of the slow AHP. The D1 receptor agonist, SKF 38393 (j), has no effect on the AHP. In total, about 35% of the tested neurons showed a prominent response to dopamine agonists with a reduced AHP or Ca^{2+} current. 5-HT as a rule had a potent effect on K_{Ca} in all neurons in the lateral cell column (compare with ref. 4). In 10 cells DA and 5-HT applied together depressed the AHP more effectively than either amine alone. Bath-application of 0.1 μ M 5-HT caused a partial reduction of the AHP (k), whereas an addition of 25 μ M DA produced a further reduction (k). Conversely, a bath application of 25 μ M DA caused a partial reduction of the AHP (l), whereas an addition of 0.1 μ M 5-HT further reduced the AHP. The frequency of locomotor burst activity (histogram in m), recorded in ventral roots (see ref. 8), induced by 100 μ M NMDA (m1, control), showed a reduction when 5-HT (m2, 0.1 μ M) was included in the perfusion solution as did DA (m3, 50 μ M). When both

compounds were applied together (m4, 50 μ M DA + 0.1 μ M 5-HT) a more marked reduction occurred. The data are representative of each of 3 experiments. Thin bars indicate s.d. during 30 burst cycles.

METHODS. Electrophysiology: Pieces of spinal cord (*Lampetra fluviatilis* and *Ichthyomyzon unicuspis*) were prepared for *in vitro* intracellular recording as described previously⁴. DA (10 mM; Sigma), apomorphine, quinpirole and SKF 38393 (200 μ M; Research Biochemicals) were applied by pressure ejection⁴. Bupropion (100–200 μ M; Research Biochemicals), citalopram (2 μ M; Lundbeck), and in some experiments DA and 5-HT (see above) were added to the perfusion solution. Recordings were stored and analysed using Clampex and AxoGraph software (Axon Instruments). AHP amplitudes were calculated as the difference between the average resting membrane potential (for the first 5 ms of the trial) and the peak amplitude of the hyperpolarizing potential. The DA-induced mean reduction in AHP amplitude of 2.22 mV was statistically significant ($P < 0.01$; paired *t*-test). Single electrode voltage clamp recordings were done with an Axoclamp 2A switch amplifier, using electrodes with 4 M CsCl, with a resistance of 13–20 M Ω . HPLC analysis of spinal cord perfusates: Pieces of spinal cord (20–30 mm) were incubated for 48 h in Ringer solution containing tropolone (100 μ M) and pargyline (1 mM) to prevent oxidation of DA^{21–23}. The HPLC analysis was as described previously¹⁶. The DA level in the control samples was 1.73 ± 0.72 nM (mean $\pm$ s.d., 5 paired samples were analysed for each condition), whereas in the presence of bupropion it was 3.13 ± 0.70 nM (statistically significant difference, $P < 0.001$, Student's *t*-test). The level of noradrenaline could not be reliably measured, but noradrenaline (20–50 μ M) does not affect the AHP (M.W. and S.G., unpublished observations, compare with ref. 7). The concentration of the DA metabolites 3,4-dihydroxyphenyl acetic acid and homovanillic acid did not differ significantly between control (3.79 ± 1.13 and 1.48 ± 0.61 nM, respectively) and bupropion samples (4.70 ± 0.79 and 1.91 ± 0.52 nM, respectively). The concentration of 5-HT was always below the detectable level (compare ref. 13), whereas that of 5-hydroxyindoleacetic acid was similar under both conditions (11.03 ± 3.84 nM and 12.66 ± 4.92 , respectively).

primary antibodies against 5-HT (Fig. 1a) and dopamine (DA; Fig. 1b), respectively (see legend). Three cells below the central canal and the characteristic dense 5-HT plexus in the ventromedial aspect of the cord can be seen to be immunoreactive towards 5-HT (Fig. 1a). In addition, 5-HT-immunoreactive fibres occur in the dorsal column and the lateral aspect of the spinal cord. These fibres are known to originate from 5-HT neurons in the brainstem and in spinal ganglia^{6,9}. In Fig. 1b the same three cells show DA immunoreactivity, as do varicose fibres in an area overlapping with the 5-HT plexus. DA immunoreactivity is also shown by small liquor-contacting cells around the central canal^{10,11}, which were not stained by the 5-HT antiserum. Conversely, the 5-HT-immunoreactive fibres in the dorsal column and dorsolateral aspect of the spinal cord did not display DA immunoreactivity. Figure 1c shows the detailed morphology of DA-immunoreactive neurons visualized with peroxidase anti-peroxidase (PAP) methodology (see legend) around the ventromedial portion of the spinal cord and in the lateral cell column (Fig. 1d).

To test the cellular effect of DA, neurons known to have dendrites in the 5-HT-DA plexus were recorded intracellularly. The isolated thin and flat lamprey spinal cord (dorsoventral thickness of 200–300 μm) was maintained *in vitro* in physiological solution⁴. Neurons in the lateral cell column were impaled with KCl-containing microelectrodes. A DA-containing extracellular pipette (10 mM) was positioned close to the spinal cord at the level of the 5-HT-DA plexus and, by application of pressure, small amounts of DA were ejected. Figure 2a shows a recording of the postspike afterhyperpolarization (AHP) in a spinal cell. The application of DA caused a marked reduction of the AHP. This reduction of the AHP could be due to either a direct action on K_{Ca} channels, as is the case with 5-HT⁴, or to a reduction of the Ca^{2+} entry during the action potential.

In Fig. 2b voltage-dependent K^{+} channels have been blocked by application of tetraethylammonium (TEA), which prevents a rapid repolarization, resulting in a marked prolongation of the action potential due to calcium channels remaining open for a longer period¹². When DA was applied, the calcium component of the action potential was reduced in parallel with the AHP (Fig. 2b). Under similar conditions, barium was used to replace Ca^{2+} in the physiological solution (Fig. 2c). Ba^{2+} ions pass through Ca^{2+} channels but, in contrast to Ca^{2+} ions, do not activate K_{Ca} channels. The long-lasting Ba^{2+} component of the action potential was markedly reduced during application of DA. To test directly the effect of DA on the Ca^{2+} current, voltage-clamp experiments were done. Figure 2d shows that the Ca^{2+} current evoked by a depolarizing step after a blockade of Na^{+} and K^{+} currents (see legend to Fig. 2) was markedly reduced by application of DA. Application of 5-HT (1–10 μM) has no effect on the Ca^{2+} current in these neurons (ref. 4 and M.W. and S.G., unpublished observations).

The 5-HT re-uptake blocker citalopram increases the extracellular level of endogenous 5-HT, and mimics the effect of exogenously applied 5-HT on the locomotor activity^{13,14}. Similarly, citalopram (2 μM) was found to mimic the effect of 5-HT on the postspike AHP (Fig. 2e, see legend). If DA is also released, the administration of a DA re-uptake blocker should elevate the extracellular levels of endogenously released DA, resulting in a reduction of the AHP. Application of the selective DA re-uptake blocker bupropion¹⁵ (100–200 μM) reduced the AHP (Fig. 2f), and the Ca^{2+} current under voltage clamp (Fig. 2g), indicating that DA is released in the spinal cord and exerts the same effect as exogenously administered DA. To ascertain that bupropion actually causes a dopamine release in lamprey, spinal cord pieces were incubated in Ringer solution with and without 100 μM bupropion, and the supernatant was used for determination of DA content by high-performance liquid chromatography (HPLC) analysis¹⁶. The DA concentration was significantly higher in the presence of bupropion than in the control solution (see legend of Fig. 2), corroborating the interpretation that the

reduction of the Ca^{2+} current observed with bupropion is due to DA.

A reduction of the AHP could also be induced by puff application of a 10 mM bolus of either the DA D1/D2 agonist apomorphine (Fig. 2h), or the D2/D3 receptor agonist quinpirole (Fig. 2i), but not by the D1 receptor agonist SKF 38393 (Fig. 2j). DA had little or no effect when the D2 receptor antagonist eticlopride (10 μM) was added to the solution (not illustrated). It thus appears that these DA effects are exerted on spinal neurons through D2 receptors.

To test whether DA and 5-HT together can depress the AHP more effectively than either amine alone, the effect of a coapplication was examined by administering 5-HT followed by DA and 5-HT or vice versa. Bath application of a low dose of 5-HT (Fig. 2k; 0.1 μM) reduced the AHP with $11.8 \pm 10.7\%$ ($\pm$ s.d., mean reduction in 5 cells), whereas the addition of DA (Fig. 2k; 25 μM) caused a further mean reduction of $39.2 \pm 14.6\%$. Conversely, an initial application of a low concentration of DA (Fig. 2l; 25 μM) caused a mean reduction of the AHP of $35.1 \pm 15.6\%$ ($n=5$), whereas the addition of 5-HT (Fig. 2l; 0.1 μM) caused a further mean reduction of $25.5 \pm 16.7\%$. The effects of DA and 5-HT on the AHP are thus additive, although their action is achieved by different mechanisms.

A DA-induced reduction in the AHP during excitatory amino-acid-induced locomotion in the lamprey would be expected to increase the intensity and duration of ventral root burst activity, as well as to slow down the cycle frequency^{7,8}. When DA (50–100 μM) was included in the perfusion solution, the duration of ventral root bursting and cycle duration was significantly increased in 7 of the 8 preparations tested ($P < 0.05$). Low doses of 5-HT (0.1–1 μM) increased the effect of DA (25–100 μM) and vice versa (Fig. 2m). Thus, both at the cellular level and at the network level, the combined action of the two transmitters exceeds that of each transmitter alone. This is in contrast to the neuromuscular junction of *Aplysia*, in which one transmitter appears to counteract the effect of its cotransmitter¹⁷.

In conclusion, 5-HT and DA act as modulators of spinal neurons controlling locomotion. DA released from 5-HT-DA midline neurons exerts a complementary action to 5-HT on the AHP of spinal neurons; DA reduces the Ca^{2+} entry during the action potential, whereas 5-HT acts directly on the K_{Ca} channels responsible for the AHP. The effects of DA may also involve other cellular mechanisms dependent on Ca^{2+} entry through ionic channels. $\square$

Received 14 November 1994; accepted 30 January 1995.

- Cuello, A. C. (ed.) *Co-transmission* (Macmillan, London, 1982).
- Hökfelt, T. in *Synaptic Function* (ed.) Edelman, G. M. 179–211 (Wiley, New York, 1987).
- Schwandt, P. C., Spain, W. J. & Crill, W. E. *J. Neurophysiol.* **67**, 216–226 (1992).
- Wallén, P. *et al. J. Neurophysiol.* **61**, 759–768 (1989).
- Harris-Warrick, R. M., McPhee, J. C. & Filler, J. A. *Neuroscience* **14**, 1127–1140 (1985).
- Van Dongen, P. A. M. *et al. J. comp. Neurol.* **234**, 501–522 (1985).
- Harris-Warrick, R. M. & Cohen, A. H. *J. exp. Biol.* **116**, 27–46 (1985).
- Schotland, J. L. & Grillner, S. *Exp. Brain Res.* **93**, 391–398 (1993).
- Brodin, L., Buchanan, J. T., Hökfelt, T., Grillner, S. & Verhofstad, A. A. *J. Neurosci. Lett.* **67**, 53–57 (1986).
- Brodin, L., Dale, N., Christenson, J., Storm-Mathisen, J., Hökfelt, T. & Grillner, S. *Brain Res.* **508**, 172–175 (1990).
- Vigh, B., Vigh-Teichmann, I., Manzano, E., Silva, M. J. & Vanden Pol, A. N. *Cell Tissue Res.* **231**, 615–621 (1983).
- Hill, R. H., Arhem, P. & Grillner, S. *Brain Res.* **358**, 40–52 (1985).
- Christenson, J., Franck, J. & Grillner, S. *Neurosci. Lett.* **100**, 188–192 (1989).
- Matsushima, T., Grillner, S. *J. Neurophysiol.* **67**, 1683–1690 (1992).
- Nomikos, G. G., Damsma, G., Wenkstern, D. & Fibiger, H. C. *Neuropsychopharmacology* **7**, 7–14 (1992).
- Herrera-Marschitz, M. *et al. Amino Acids* **2**, 157–179 (1992).
- Church, P. J., Whim, M. D. & Lloyd, P. E. *J. Neurosci.* **13**, 2790–2800 (1993).
- Steinbusch, H. W. M., Van Vliet, S. P., Bol, J. G. J. M. & DeVente, J. in *Neurocytochemical Methods* (eds Culas, A. & Eugène, D.) NATO ASI series, Vol. 58 (Springer, Berlin, 1991).
- Verhofstad, A. A. J. *et al. Cytochemistry: Practical Applications in Pathology and Biology* (eds Polak, J. M. & Van Norden, S.), (Wright, Bristol, 1983).
- Van Dongen, P. A. M., Grillner, S. & Hökfelt, T. *Brain Res.* **366**, 320–325 (1986).
- Dedek, J., Baumes, R., Tien-Duc, N., Gomeni, R. & Korf, J. *J. Neurochem.* **33**, 687–695 (1979).
- Sharp, T., Zetterström, T. & Ungerstedt, U. *J. Neurochem.* **47**, 113–122 (1986).
- Cumming, P., Brown, E., Damsma, G. & Fibiger, H. J. *J. Neurochem.* **59**, 1905–1914 (1992).

ACKNOWLEDGEMENTS. We thank H.W.M. Steinbusch and A.A.J. Verhofstad for the generous supply of DA and 5-HT antisera. This work was supported by Fogarty and IBRO Fellowships to J.S. and the Swiss National Foundation for Science to W.Z. and by the Swedish Medical Research Council.

Malarial haemozoin/ β -haematin supports haem polymerization in the absence of protein

Arnulf Dorn*, Ruedi Stoffel*, Hugues Matile*,
André Bubendorf† & Robert G. Ridley*

Pharma Division, Preclinical Research in * Infectious Diseases and
† Spectroscopy, F. Hoffmann-La Roche Ltd, CH-4002 Basel,
Switzerland

MALARIAL parasites growing inside erythrocytes digest up to 80% of the host cell's haemoglobin within a lysosomal organelle, the digestive vacuole^{1,2}. They sequester the potentially toxic haem (Fe (II) protohaematoporphyrin) that is released during this process into an insoluble pigment called haemozoin, which consists of polymerized Fe (III) protohaematoporphyrin subunits³. We have studied this process of haem polymerization, which was previously reported to be enzyme-mediated and the target of the quinoline antimalarial drugs chloroquine and quinine⁴. Here we show that, rather than being enzyme-mediated, haem polymerization is actually a chemical process, dependent only on the presence of haem-derived material associated with haemozoin and not on protein. This discovery does not invalidate haem polymerization as a target for drug intervention and the mechanism by which haemozoin formation is initiated is still not understood, but our view of this process and of the action of chloroquine must be reconsidered.

In the assay originally used to describe 'haem polymerase' activity⁴, radiolabelled ¹⁴C-haem was added to a parasite lysate at acid pH (pH 4.8; equivalent to the pH of the digestive vacuole) and the incorporation of radioactivity into insoluble haemozoin was measured. The activity was reported to be labile, destroyed by SDS, and associated with a membrane-pellet fraction, suggesting that it was attributable to a membrane-associated enzyme of the food vacuole⁴. It was also reported that this 'haem polymerase' activity could be extracted from parasite lysate using acetonitrile⁵.

We investigated the lability of this 'haem polymerase' activity and found that parasite lysates retained their activity at 4 °C for several weeks and even after boiling (Fig. 1), raising doubts as to whether the 'haem polymerase' activity was indeed protein-mediated.

To clarify the mechanism of haem polymerization in the

parasite, we confirmed that the activity was associated with the insoluble fraction of the parasite lysate, which contains the digestive vacuoles⁴ (Fig. 2a); crude digestive vacuole preparations were then subfractionated^{6,7} into membrane, digestive-vacuole and haemozoin fractions, which were tested for their ability to support haem polymerization (Fig. 2b). The results showed an enrichment of 'haem polymerase' activity in the haemozoin fraction; activity in the membrane and vacuole fractions could be accounted for by the presence of haemozoin pigment in these fractions. Treatment of haemozoin with proteinase K and chloroform-methanol^{8,9} to remove associated proteins (confirmed by silver-stained SDS-PAGE analysis on 20% polyacrylamide) gave material with activity comparable to untreated haemozoin (Fig. 2c). Exposure of parasite lysate to a variety of proteinases (proteinase K, pronase E, Lys C and pepsin) also had no effect on haem polymerization.

These results suggested that haemozoin is sufficient to support 'haem polymerase' activity. We also tested the ability of β -haematin to promote haem polymerization (Fig. 3) as its structure is similar to that of haemozoin³; it is prepared by heating a haem chloride solution to 70 °C at acid pH and so contains no protein. We found that it could support haem polymerization at 37 °C like haemozoin and that its activity could be inhibited by chloroquine (Fig. 3a).

The reported inhibition of haem polymerization by SDS⁴ also occurred when polymerization was promoted by β -haematin (Fig. 3b), which suggests that inhibition by SDS is not a result of protein denaturation but of interference with the assembly or structure of the β -haematin/haemozoin polymer. This idea is supported by the observation that the dimerization of Fe(III) porphyrins via μ -oxo bridges is inhibited when haemin is solubilized in SDS micelles¹⁰.

Another result implicating the existence of a haem polymerase enzyme still required explanation. This was the report of a heat-labile 'haem polymerase' activity extracted from trophozoite lysates using acetonitrile⁵. We confirmed that a soluble 'haem polymerase' activity can be extracted from trophozoite lysate, but found no evidence that this activity was labile: it was resistant to boiling and to pronase E (results not shown) and no protein material was detectable.

We reasoned that if this activity was due to haem-derived material, rather than to a protein, then it should be possible to extract a solubilized haem-polymerization activity from synthetic β -haematin. This was confirmed by the experiments summarized in Fig. 4, which show that acetonitrile extracts from synthetic β -haematin, trophozoite lysate and haemozoin can all support haem polymerization equally, and that they are all inhibited by

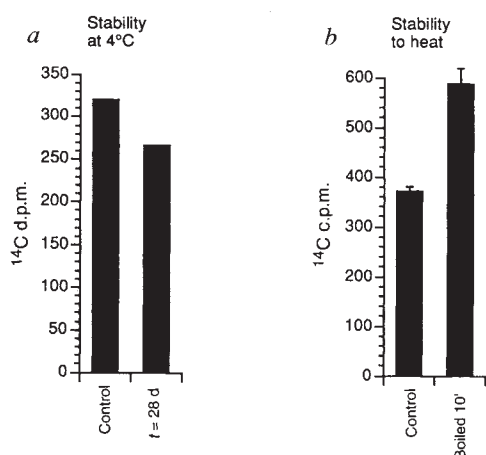


FIG. 1 Stability of haem polymerization activity in malarial trophozoite lysate. *a*, *Plasmodium falciparum* K1 trophozoite lysate was stored at 4 °C for 4 weeks ($t=28$ d) before testing it for haem polymerization activity. *b*, *Plasmodium falciparum* K1 trophozoite lysate was preincubated for 10 min at either 37 °C (control) or 100 °C and then tested for haem polymerization activity in 500 mM sodium acetate, pH 4.8, containing 140 μ M ¹⁴C-haematin. Results (mean $\pm$ s.d.) of triplicate assays are shown.

METHODS. *P. falciparum* strain K1 was cultured in A⁺ human erythrocytes according to ref. 13, modified using lipid-rich BSA (AlbuMAX I, Life Technologies, Gibco BRL) as a substitute for human serum. The culture medium consisted of RPMI 1640 supplemented with 25 mM HEPES, 50 μ g ml⁻¹ hypoxanthine, 0.21% (w/v) NaHCO₃, 50 μ g ml⁻¹ neomycin sulphate and 0.5% (w/v) AlbuMAX I. The parasite cultures were grown at 37 °C in 5% CO₂, 3% O₂, 92% N₂. Synchronization was attained by treatment with sorbitol¹⁴. Trophozoites were collected after saponin lysis¹⁵, washed in PBS, pH 7.0, and resuspended in PBS. They were then frozen as beads in liquid nitrogen and stored at -80 °C. Trophozoite lysate was prepared by thawing the beads and mechanically disrupting them with a Dounce homogenizer. Haem polymerization was assayed as previously described, after overnight incubation at 37 °C in 500 mM sodium acetate, pH 4.8, in the presence of 140 μ M ¹⁴C-haematin⁴.

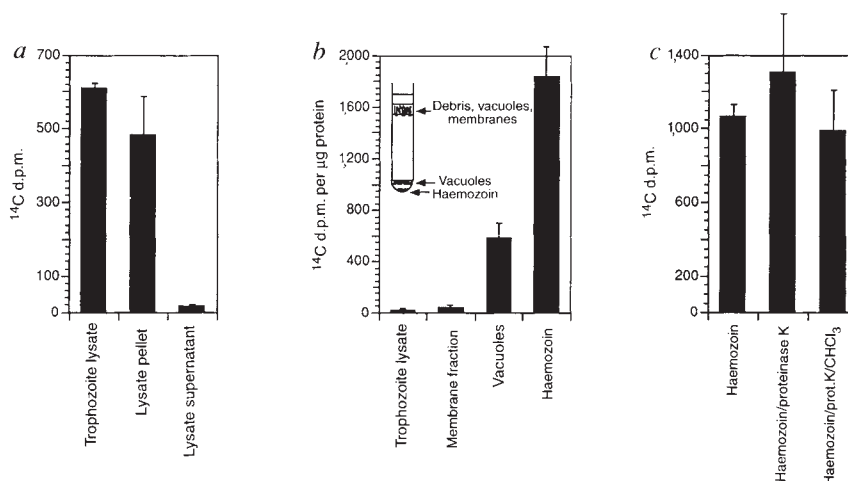


FIG. 2 Determining the source of haem polymerization activity in malarial trophozoite lysate. *a*, Activity of insoluble and soluble fractions of trophozoite lysate. Results (mean $\pm$ s.d.) of triplicate assays are shown. *b*, Activity of membrane fraction, pure vacuoles and haemozoin obtained after Percoll density-gradient fractionation⁷ of parasite lysate. Activities are normalized on the basis of protein content. Results (mean $\pm$ s.d.) of duplicate assays are shown. *c*, Activity of purified haemozoin⁶, treated with proteinase K, with and without chloroform-methanol extraction⁹. Results (mean $\pm$ s.d.) of duplicate assays are shown.

METHODS. For *a*, an aliquot of *P. falciparum* K1 trophozoite lysate (25 μg protein), obtained as described for Fig. 1, was pelleted for 15 min in an Eppendorf centrifuge. The supernatant was decanted, the pellet resuspended in 10 mM bis-Tris, pH 7.0, and both fractions tested for haem polymerization activity. For *b*, further fractionation was carried out as described⁷. Three main fractions were obtained: top, a band of membranous debris and vacuoles; close to the bottom, a band of pure

vacuoles; and at the bottom, a black pellet containing malaria pigment (haemozoin). These fractions were diluted with 5 volumes of 0.25 M sucrose, 10 mM sodium phosphate, 1.5 mM MgCl_2 , pH 7.1, centrifuged at 16,000g for 15 min and resuspended in the same buffer⁷ before testing for haem polymerization. For *c*, haemozoin was purified as in ref. 6, washed in PBS, and resuspended in 0.9 M acetic acid. The material was stored at -20°C and briefly sonicated before use to resuspend haemozoin. For proteinase K treatment, haemozoin was pelleted in an Eppendorf centrifuge for 15 min, washed and resuspended in 85 μl water. To this was added 5 μl proteinase K (50 μg μl^{-1}) and 10 μl 10 $\times$ proteinase-K buffer (Boehringer); the mixture was incubated overnight at 37°C . The haemozoin was collected by centrifugation, washed three times with 1 ml 0.1 M Tris, 0.01 M EDTA, 0.5% Triton X-100, pH 8.3, and redissolved in 150 μl distilled water. Methanol-chloroform extraction was carried out as described⁹ and haem polymerization measured as before⁴.

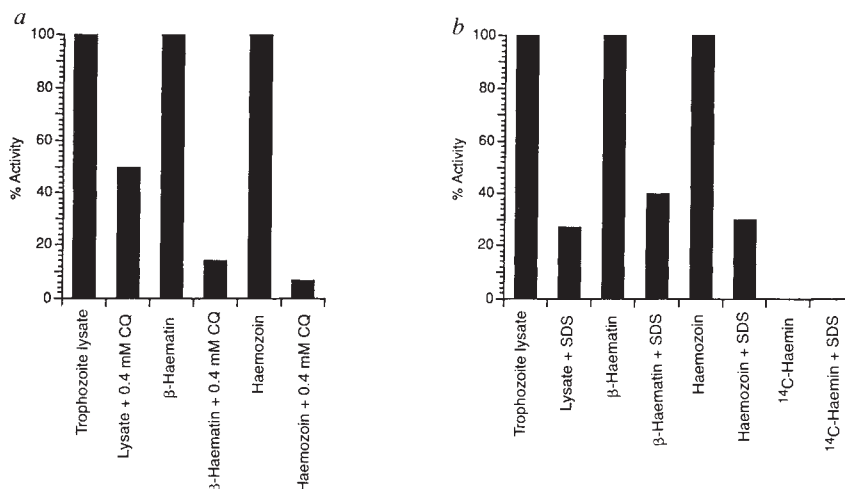


FIG. 3 β -Haematin promotes haem polymerization in the same way as haemozoin. *a*, Haem polymerization assayed with *P. falciparum* K1 lysate, β -haematin and purified haemozoin and the inhibitory effect of chloroquine (CQ). *b*, Effect of SDS on haem polymerization assayed using the same samples.

METHODS. Trophozoite lysates, haemozoin and ^{14}C -haemin were prepared as described for Fig. 1; haemozoin used here was not treated with proteinase K. β -Haematin was prepared according to ref. 3, washed and resuspended in distilled water and stored in aliquots at -20°C . *a*, Trophozoite lysate (25 μg protein), β -haematin and haemozoin were

incubated at 37°C in 500 mM sodium acetate, pH 4.8, overnight in the presence of 140 μM ^{14}C -haemin, both with and without 400 μM chloroquine. *b*, Trophozoite lysate (25 μg protein), β -haematin, haemozoin and ^{14}C -haemin (control) were preincubated at 37°C for 1 h in 2% SDS. After preincubation, samples were incubated overnight at 37°C in 500 mM sodium acetate, pH 4.8, in the presence of 140 μM ^{14}C -haemin and a final concentration of 0.1% SDS. Results are expressed as per cent inhibition, relative to haemozoin formation in the SDS-free control, and are means of duplicate assays.

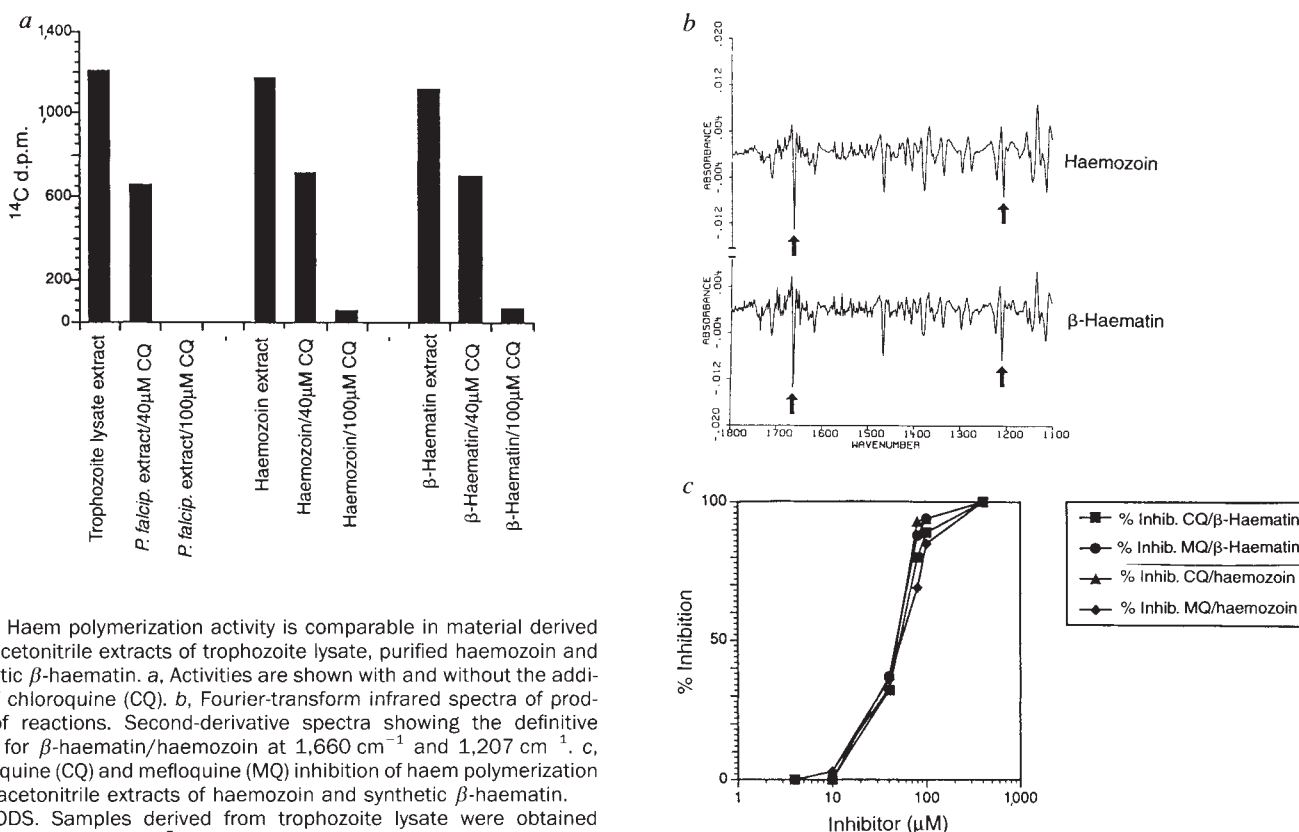


FIG. 4 Haem polymerization activity is comparable in material derived from acetonitrile extracts of trophozoite lysate, purified haemozoin and synthetic β -haematin. *a*, Activities are shown with and without the addition of chloroquine (CQ). *b*, Fourier-transform infrared spectra of products of reactions. Second-derivative spectra showing the definitive peaks for β -haematin/haemozoin at $1,660\text{ cm}^{-1}$ and $1,207\text{ cm}^{-1}$. *c*, Chloroquine (CQ) and mefloquine (MQ) inhibition of haem polymerization using acetonitrile extracts of haemozoin and synthetic β -haematin.

METHODS. Samples derived from trophozoite lysate were obtained essentially as described⁵, with minor variations. *P. falciparum* trophozoites, strain K1, were homogenized in a Dounce homogenizer in 20 mM bis-Tris-propane, pH 7.3 (buffer A) in the presence of several protease inhibitors (1 mM Pefabloc SC (4-(2-aminoethyl)-benzenesulphonylfluoride) (AEBSF); 0.1 mM leupeptin; $1\text{ }\mu\text{g ml}^{-1}$ pepstatin) (Boehringer Mannheim) and centrifuged at 25,000g in an SS34 Sorvall rotor for 40 min at 4 °C. The supernatant was removed and the pellet resuspended by brief sonication in buffer A containing 1 M NaCl and 10% acetonitrile. The sonicate was diluted in three volumes of acetonitrile, shaken for 3 h at 4 °C and then centrifuged. The upper phase of the supernatant was collected and the pellet resuspended in 2 ml buffer A, 1 M NaCl, 10% acetonitrile. This mixture was shaken at 4 °C for 3 h, left overnight at 4 °C and centrifuged at 18,000 r.p.m. in a Sorvall SS34 rotor for 40 min at 4 °C. The soluble upper phases were pooled, diluted 1:10 in buffer A to give an acetonitrile concentration of 10%, and loaded on to a Resource-Q column (1 ml; Pharmacia). Bound material

was eluted with a linear gradient from 0–1 M NaCl (buffer A plus 10% acetonitrile). The fractions were tested for haem polymerization activity and positive fractions were used in the experiment. Samples derived from haemozoin and synthetic β -haematin were obtained as follows. Purified haemozoin⁶ and synthetic β -haematin³ were pelleted and resuspended by brief sonication in buffer A containing 1 M NaCl, 10% acetonitrile. Three volumes of acetonitrile were then added and the mixture shaken for 3 h at 4 °C. The mixture was left overnight at 4 °C, centrifuged, the upper phase collected and the extraction procedure repeated. The upper phases were pooled, diluted 1:4 in buffer A and tested for haem polymerization activity. Fourier transform infrared spectra are recorded as KBr pellets, with a Nicolet 20SXB FTIR spectrometer. A sum of 512 scans are accumulated at a resolution of 2 cm^{-1} . The comparison is achieved in the spectral region $1,800\text{--}1,100\text{ cm}^{-1}$ as transmission spectra and as their second derivatives.

was eluted with a linear gradient from 0–1 M NaCl (buffer A plus 10% acetonitrile). The fractions were tested for haem polymerization activity and positive fractions were used in the experiment. Samples derived from haemozoin and synthetic β -haematin were obtained as follows. Purified haemozoin⁶ and synthetic β -haematin³ were pelleted and resuspended by brief sonication in buffer A containing 1 M NaCl, 10% acetonitrile. Three volumes of acetonitrile were then added and the mixture shaken for 3 h at 4 °C. The mixture was left overnight at 4 °C, centrifuged, the upper phase collected and the extraction procedure repeated. The upper phases were pooled, diluted 1:4 in buffer A and tested for haem polymerization activity. Fourier transform infrared spectra are recorded as KBr pellets, with a Nicolet 20SXB FTIR spectrometer. A sum of 512 scans are accumulated at a resolution of 2 cm^{-1} . The comparison is achieved in the spectral region $1,800\text{--}1,100\text{ cm}^{-1}$ as transmission spectra and as their second derivatives.

We conclude that the 'haem polymerase' activity observed in parasite lysates is due to an autocatalytic chemical process of haem polymerization on to haem-derived material associated with haemozoin, and is not due to enzyme catalysis. This does not invalidate the important ideas that haem polymerization is the target of chloroquine⁴ and that this process could be a target for new drugs. Also, as our results apply only to the elongation stage of haem polymerization, a protein could still be instrumental in initiating the whole process. If such a protein exists, however, we envisage that it will function not as a 'haem polymerase', but rather as a structural focus about which the initial oligomerization of haem can occur. □

Received 11 November 1994; accepted 31 January 1995.

- Goldberg, D. E. & Slater, A. F. G. *Parasit. Today* **8**, 280–283 (1992).
- Slater, A. F. G. *Pharmac. Ther.* **57**, 203–235 (1993).
- Slater, A. F. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 325–329 (1991).
- Slater, A. F. G. & Cerami, A. *Nature* **355**, 167–169 (1992).
- Slater, A. F. G. & Cerami, A. *Int. Patent Appl. PCT/US92/11279* (1993).
- Ashong, J. O., Blench, I. P. & Warhurst, D. C. *Trans. R. Soc. trop. Med. Hyg.* **83**, 167–172 (1989).
- Goldberg, D. E., Slater, A. F. G., Cerami, A. & Henderson, G. B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2931–2935 (1990).
- Fitch, C. D. & Kanjanangkul, P. *J. biol. Chem.* **262**, 15552–15555 (1987).
- Cohen, P., Broekman, M. J., Verkley, A., Lisman, J. W. W. & Derksen, A. *J. clin. Invest.* **50**, 762–772 (1971).

- Gushimana, Y., Doepner, B., Martinez-Hackert, E. & Ilgenfritz, G. *Biophys. Chem.* **47**, 153–162 (1993).
- Egan, T. J., Ross, D. C. & Adams, P. A. *FEBS Lett.* **352**, 54–57 (1994).
- Fitch, C. D. *et al. Antimicrob. Agents Chemother.* **21**, 819–822 (1982).
- Trager, W. & Jensen, J. B. *Science* **193**, 673–675 (1976).
- Lambros, C. & Vanderberg, J. P. *J. Parasit.* **65**, 418–420 (1979).
- Goman, M. *et al. Molec. Biochem. Parasit.* **5**, 391–400 (1982).

ACKNOWLEDGEMENTS. We thank M. Page, W. Hofeinz, W. Leupin and R. Moon for discussion and for critically reading the manuscript, and M. Kleinmann for technical assistance.

Physiological effects of inverse agonists in transgenic mice with myocardial overexpression of the β_2 -adrenoceptor

Richard A. Bond*, Paul Leff†, T. David Johnson‡, Carmelo A. Milano§, Howard A. Rockman||, Thomas R. McMin||, Subramaniam Apparsundaram*, Michael F. Hyek*, Terry P. Kenakin¶, Lee F. Allen#** & Robert J. Lefkowitz*

* Department of Pharmacological and Pharmaceutical Sciences, University of Houston, Houston, Texas 77204, USA

† Department of Pharmacology, Fisons plc, Loughborough LE11 0RH, UK

‡ Department of Anesthesiology, Baylor College of Medicine, Houston, Texas 77030, USA

§ Departments of § Surgery, ☆ Cardiology and # Hematology/Oncology

and the ☆ Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710, USA

|| Department of Medicine, University of California at San Diego, School of Medicine, La Jolla, California 92093, USA

¶ Glaxo Research, Research Triangle Park, North Carolina 27709, USA

USA

G-PROTEIN-COUPLED receptors are thought to have an inactive conformation (R), requiring an agonist-induced conformational change for receptor/G-protein coupling¹⁻³. But new evidence suggests a two-state model⁴⁻¹⁹ in which receptors are in equilibrium between the inactive conformation (R), and a spontaneously active conformation (R*) that can couple to G protein in the absence of ligand (Fig. 1). Classic agonists have a high affinity for R* and increase the concentration of R*, whereas inverse agonists have a high affinity for R and decrease the concentration of R*. Neutral competitive antagonists have equal affinity for R and R* and do not displace the equilibrium, but can competitively antagonize the effects both of agonists and of inverse agonists. The lack of suitable *in vivo* model systems has restricted the evidence for the existence of inverse agonists to computer simulations^{7,8} and *in vitro* systems^{5,9-12,20-23}. We have used a transgenic mouse model in which there is such marked myocardial overexpression of β_2 -adrenoceptors that a significant population of spontaneously activated receptor (R*) is present, inducing a maximal response without agonist²⁴. We show that the β_2 -adrenoceptor ligand ICI-118,551 functions as an inverse agonist, providing evidence supporting the existence of inverse agonists and validating the two-state model of G-protein-coupled receptor activation.

To assess inverse agonist effects, three lines of transgenic mice with cardiac-specific overexpression of the wild-type β_2 -adrenoceptor²⁴ and control (non-transgenic) animals were investigated. Baseline atrial tension, assessed by isometric tension development in isolated left atria, was increased roughly threefold in transgenic lines with marked overexpression of the wild-type β_2 -adrenoceptor (TG-4, TG-33)²⁴. ICI-118,551, a β_2 -adrenoceptor antagonist, inhibited baseline left atrial tension in these animals by 20–80% and the inhibitory effect correlated with β_2 -adrenoceptor densities, that is, lines with higher receptor densities showed more inhibition (data not shown). This suggests that the inhibitory response to ICI-118,551 was receptor mediated and not a nonspecific event.

In atria from control mice, isoprenaline produced concentration-dependent increases in left atrial tension (Fig. 2a). The half-maximal effective concentration (EC_{50}) for isoprenaline in control mice (1.8×10^{-10} M) was shifted only fourfold to the right (7.7×10^{-10} M) by ICI-118,551, but 2,000-fold (3.9×10^{-7} M) in the presence of both ICI-118,551 and the β_1 -adrenoceptor antagonist, CGP-20712A; the change in slope of the isoprenaline-response curve with ICI-118,551 results from the antagonist-induced conversion from a two-receptor- (β_1 and β_2 -adrenoceptors) to a one-receptor- (β_2 -adrenoceptor) mediated process. In transgenic animals, atrial tension is maximally stimulated at baseline and no response to isoprenaline was observed (Fig. 2b). Following ICI-118,551-induced inhibition of left atrial tension, however, isoprenaline increased left atrial tension, producing responses that restored this parameter to pre-antagonist values (Fig. 2b). In the presence of ICI-118,551, the EC_{50} for isoprenaline in transgenic atria (3×10^{-8} M) was shifted almost 170-fold when compared with its EC_{50} in control mice (1.8×10^{-10} M). No further shift was noted after the addition of CGP-20712A, indicating that functionally the response was mediated by a homogeneous population of β_2 -adrenoceptors. To demonstrate the specificity of isoprenaline's ability to reverse the inhibitory effect of ICI-118,551, comparable reductions in baseline tension were produced using the cholinergic agonist, methacholine

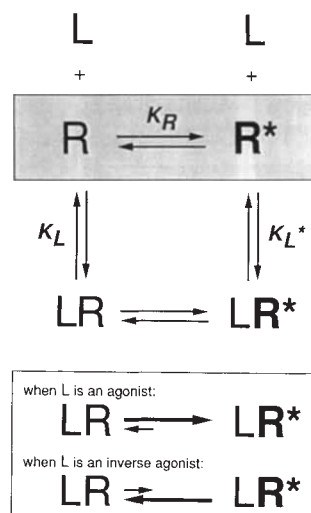


FIG. 1 The two-state model of receptor activation depicting the receptor existing in inactive (R) and active (R*) conformations; L, ligand; K_R , the equilibrium constant for the distribution of receptor between R and R* in the absence of ligand; K_L and K_{L^*} , the dissociation equilibrium constant for ligand at the two receptor states. These equilibria are defined as follows: $K_R = [R]/[R^*]$; $K_L = [R][L]/[LR]$; $K_{L^*} = [R^*][L]/[LR^*]$; and the total receptor population ($[R]_{TOT} = [R] + [R^*] + [LR] + [LR^*]$). Using these equations, a relationship can be derived between the concentration of the agonist ([L]) and the concentration of receptors in the active form ($[R^*] + [LR^*]$), which is expressed as the fraction of the total receptor concentration ($f_{R^*} = (K_L + [L])/K_L(1 + K_R) + (1 + K_R K_{L^*}/K_L)[L]$). The interactions between two ligands can be modelled by including the additional equilibrium equations, $K_B = [R][B]/[BR]$; $K_{B^*} = [R^*][B]/[BR^*]$, and extending the distribution equation¹⁵. Agonist concentration-response curves were simulated under two conditions: (1) the measured response is directly proportional to f_{R^*} , that is, with no receptor reserve; (2) the measured response is a rectangular hyperbolic function of f_{R^*} , that is, the fraction of receptors in the active state (f_{R^*}) required for a full response could be very small and therefore receptor reserve was present. In the latter case, an additional parameter, K_E is introduced, this being the value of f_{R^*} for half-maximal effect. Under this condition the difference between normal and TG-4 animals is modelled as a simple upregulation of receptor density; the absolute quantities of both R and R* increase, but the ratio remains the same. Under both conditions, the model fits the data equally well.

** Present address: Huntsman Cancer Institute, University of Utah, Health Sciences Center, Building 570, Room 410B, Salt Lake City, Utah 84112, USA.

(1 μ M), or the calcium channel blocker, verapamil (100 nM). Under these conditions, isoprenaline was unable to reverse the inhibition of atrial tension by methacholine and only partially able (<50%) to reverse the inhibition by verapamil (data not shown). Conversely, if control mice were converted to a hypercontractile state (resembling TG-4 mice) through a non- β_2 -adrenoceptor mechanism, that is, by increasing atrial tension with forskolin (10 μ M), ICI-118,551 had no inhibitory effect (data not shown). These results indicate that the mechanism of ICI-118,551's inhibition of baseline atrial tension was through decreasing β -adrenoceptor-mediated increases in contractility.

If competition with endogenous catecholamines were the mechanism of ICI-118,551's inhibition, then any antagonist with affinity for the β_2 -adrenoceptor used in the same relative concentration should produce an equal amount of inhibition. However, four different β -adrenoceptor antagonists, used at concentrations 300 times their K_B values (see Fig. 1) for the β_2 -adrenoceptor²⁵, were found to produce widely varying degrees of inhibition of left atrial tension (alprenolol, propranolol, ICI-118,551 and nadolol produced $14 \pm 6\%$, $38 \pm 5\%$, $76 \pm 3\%$ and $82 \pm 4\%$ inhibition of baseline left atrial tension, respectively, data not shown). Treatment with the catecholamine-depleting agent, reserpine (0.3 mg per kg (body weight) intraperitoneally (ip), 24 h before killing) produced a >97% decrease in cardiac catecholamine content²⁶. This dose of reserpine, however, did not affect baseline tension in the isolated atria of either TG-4 or control animals, and baseline responses in TG-4 animals remained comparable to the maximum response to isoprenaline in control animals (data not shown). In addition, the inhibitory concentration-response curve to ICI-118,551 (1×10^{-9} M to 3×10^{-6} M) was unaffected by reserpine treatment (data not shown). Thus, these experiments demonstrate that the inhibition of left atrial tension by ICI-118,551 does not result from antagonism of endogenous noradrenaline or adrenaline at the β -adrenoceptors.

The most conclusive evidence for the specificity of ICI-118,551's inhibition is provided by experiments assessing the net effect of the two β -adrenoceptor antagonists, ICI-118,551 and alprenolol. If the inhibitory effect of ICI-118,551 were due to either competition with endogenous noradrenaline or a non-specific effect, then the combination of these two antagonists should have exacerbated the effect, or at least left the larger effect intact. ICI-118,551 (1×10^{-9} M to 1×10^{-5} M) produced

a concentration-dependent inhibition of baseline tension in the atria of TG-4 mice (Fig. 3a). The concentration-response curve for ICI-118,551 was displaced to the right by 30-fold (measured at the EC_{40}) by the competitive (neutral) antagonist alprenolol (1×10^{-7} M). This shift is consistent with alprenolol's known affinity for β_2 -adrenoceptors²⁵ and is indicative of a competitive interaction between the two compounds for a single receptor site.

Experiments done in isolated atria with the β -adrenoceptor alkylating agent, pindobind, rule out a one-receptor-state model for receptor-ligand interactions. Pindobind (1×10^{-7} M for 15 min) decreased baseline tension of TG-4 atria by 25%. After pindobind treatment, the potency of ICI-118,551 increased with a 30-fold shift in the EC_{50} for ICI-118,551 to the left (Fig. 3c); this contrasts with a 30-fold decrease in the potency of isoprenaline after pindobind treatment of control animals (data not shown).

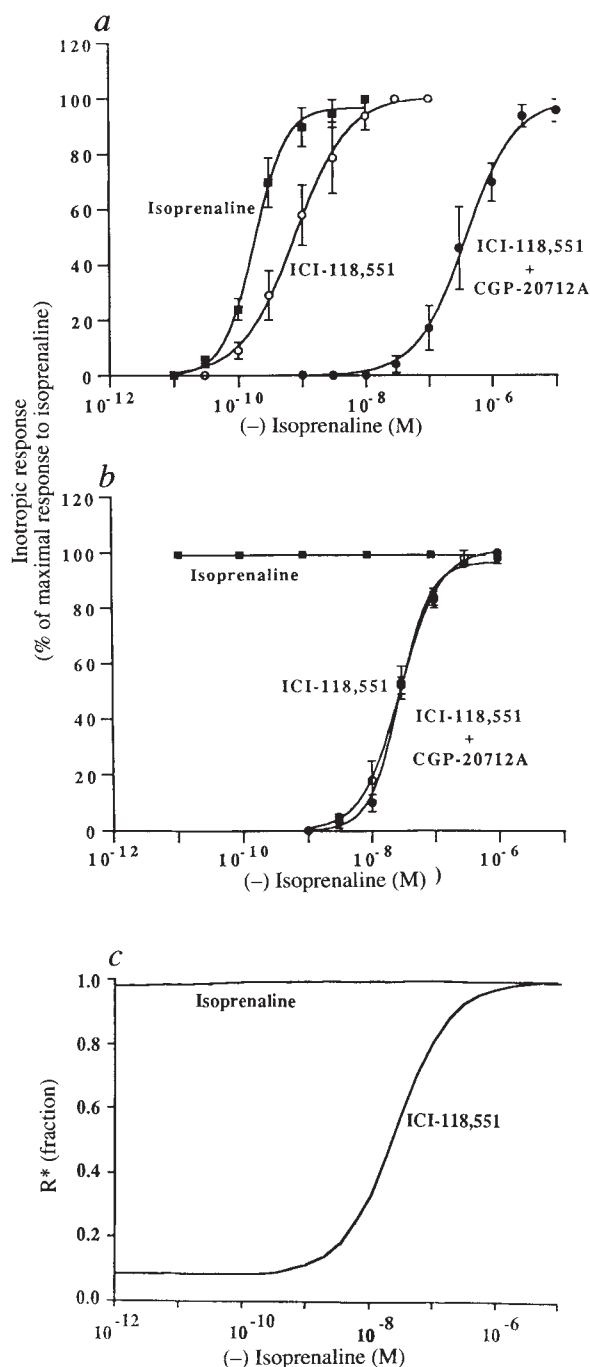


FIG. 2 The atria were suspended under optimal tension in modified Krebs's bicarbonate solution at 30 °C²⁴, and paced at the optimal frequency for tension development (3.2 Hz), 3 ms duration, voltage at threshold +20%; the results are expressed as the mean per cent inhibition $\pm$ s.e.m. ($n=4$ to 6 for each group); a, Left atrial tension development in response to isoprenaline in control and b, TG-4 mice. Isoprenaline concentration-response curve: control, $\blacksquare$; + ICI-118,551 (3×10^{-7} M), $\circ$; + ICI-118,551 (3×10^{-7} M) and CGP-20712A (3×10^{-7} M), $\bullet$. Inotropic responses are expressed as the percentage of the maximal response to isoprenaline and data reported as the mean values $\pm$ s.e.m. ($n=8$ for control and $n=5$ for TG-4); baseline values for control and TG-4 mice were 83 ± 10 mg and 262 ± 42 mg, respectively. c, Simulation based on the simple form of the two-state model (no receptor reserve) using the following parameters: K_R (basal R/R^*) = 0.02; for isoprenaline, $K_L = 1 \times 10^{-6}$ M, and $K_L^* = 2 \times 10^{-9}$ M; for ICI-118,551, $K_L = 5 \times 10^{-10}$ M, and $K_L^* = 3 \times 10^{-6}$ M. Virtually identical simulations were obtained if receptor reserve was included (see legend to Fig. 1) using the same ligand parameter values, but with $K_R = 10$ and $K_E = 0.002$.

METHODS. Transgenic mice were produced using a transgene construct incorporating the alpha myosin heavy chain promoter ligated to the human β_2 -adrenoceptor²⁴; intense, cardiac-specific, overexpression of the adrenoceptor was achieved in these transgenic animals. Three lines overexpressing the wild-type β_2 -adrenoceptor were investigated: TG-4, overexpressing the β_2 -adrenoceptor at about 200 times the total cardiac β -adrenoceptor densities of control animals; TG-33, with receptor densities about 90-fold above control; TG-35, with densities about 50-fold above control²⁴.

Using the two-state model of receptor activation (Fig. 1), the interactions between isoprenaline and ICI-118,551 (Fig. 2c), ICI-118,551 and alprenolol (Fig. 3b), and ICI-118,551 and pindobind (Fig. 3d) were simulated. This was done using two different sets of assumptions, both of which yielded essentially the same results. In one set of simulations we assumed that receptor reserve characterized the system such that only a small fraction of activated receptors (f_{R^*}) was required for full response. Under these circumstances the high basal tension is explained without change in the receptor equilibrium simply by saturation of signal transduction processes by the increased tissue levels of R^* . In the other set of simulations we assumed no receptor reserve and attributed the high basal tensions to a shift in the receptor state equilibrium towards R^* . With both sets of assumptions isoprenaline was unable to evoke a further response (control curve). In both cases, the effects of ICI-118,551 were modelled assuming that this ligand functions as an inverse agonist, displacing receptors back towards the inactive form (R). Under this condition, isoprenaline could now evoke a full response in TG-4 tissues (Fig. 2c). The simulation(s) parallels the experimental data, demonstrating that the model can accurately account for this ligand interaction. Likewise, the concen-

tration-response curve of ICI-118,551 could be simulated using the two-state model assuming alprenolol had equal affinity for the two receptor states (Fig. 3b) with good correspondence between the simulation(s) and the experimental data. Pindobind lowered baseline tension and markedly shifted the ICI-118,551 concentration-response curve to the left. These effects were simulated assuming that pindobind alkylates both receptor states equally, causing a reduction in the total receptor population without a change in the ratio. This decreases R^* to a level below that which saturates signal transduction processes and accounts for the decrease in basal tension produced by pindobind. Because there is also less functional opposition to the inverse agonist effects of ICI-118,551, its curve is left-shifted. An identical simulation could be achieved without the assumption of receptor reserve, but this involved the perhaps unnecessarily speculative assumption that pindobind is able to alter K_R . In both cases, the model provided a good fit (Fig. 3d) to the experimental data.

To assess the effect of ICI-118,551 and alprenolol on left ventricular performance *in vivo*, the maximum first derivative of the left ventricular systolic pressure ($LV\ dP/dt_{max}$), an index of cardiac contractility, was measured in TG-4 and control mice²⁴. In TG-4 animals, ICI-118,551 reduced $LV\ dP/dt_{max}$ by about

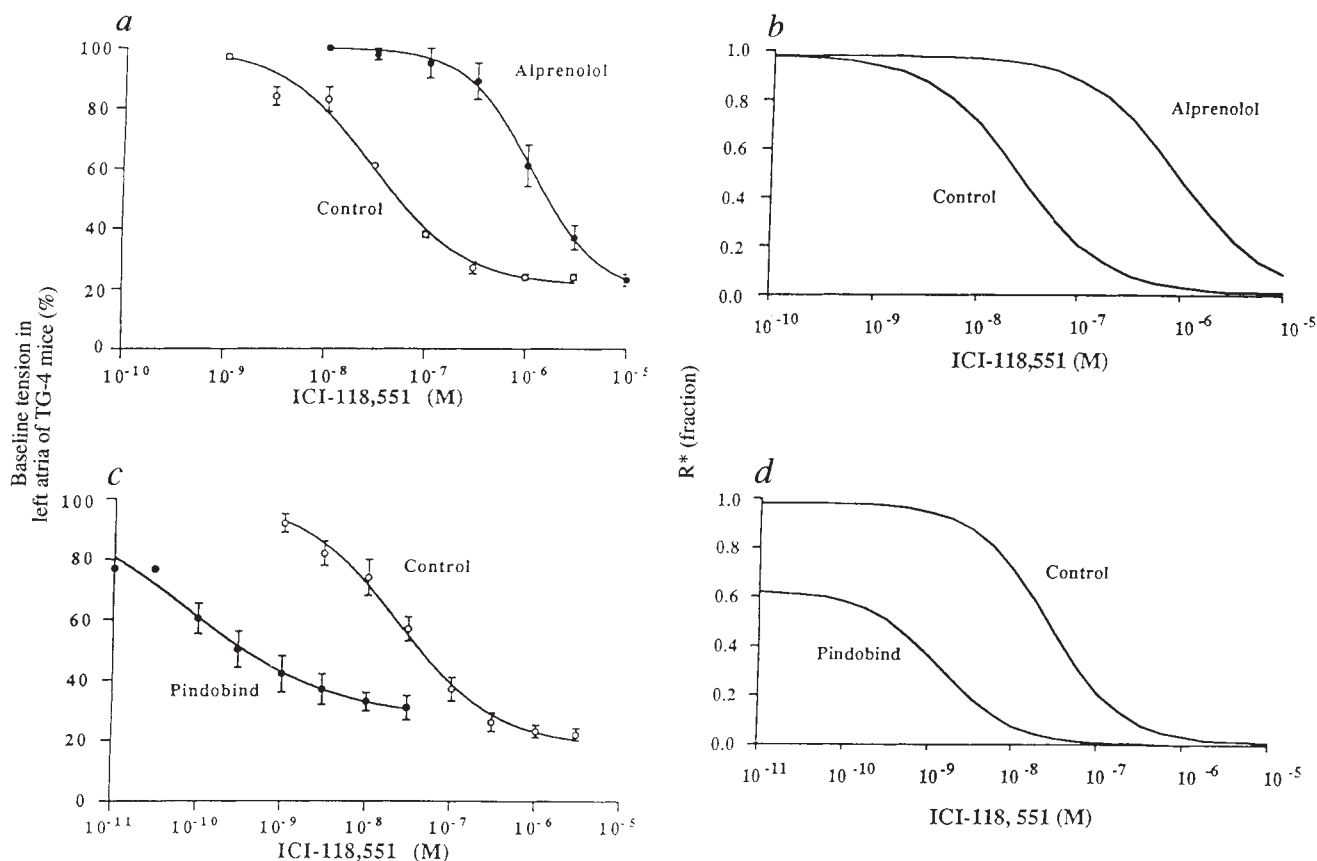


FIG. 3 a, The effect of alprenolol on the inhibitory response of ICI-118,551 on baseline left atrial tension in TG-4 mice; control, $\circ$; alprenolol (1×10^{-7} M), $\bullet$. The results are expressed as the mean per cent inhibition of baseline $\pm$ s.e.m. (baseline $\pm$ s.e.m. = 271 ± 35 mg; $n=5$ for ICI-118,551 alone; and 223 ± 40 mg; $n=5$ for ICI-118,551 in the presence of alprenolol). b, Simulation based on the simple form of the two-state model using the following parameters; K_R (basal R/R^*) = 0.02; for ICI-118,551, $K_L = 5 \times 10^{-10}$ M, and $K_L^* = 3 \times 10^{-6}$ M; alprenolol was defined as a neutral competitive antagonist, $K_L = K_L^* = 3 \times 10^{-9}$ M (based on a pA_2 ($-\log K_B$) value of 8.5 (ref. 25)). Incorporating receptor reserve (see legend to Fig. 1), virtually identical simulations were obtained using the same ligand parameters and $K_R = 10$ and $K_E = 0.002$. c, The effect of pindobind (1×10^{-7} M, 15 min incubation) on the inhibitory response of ICI-118,551 on baseline left atrial tension in

TG-4 mice; control, $\circ$; pindobind, $\bullet$. d, Simulation based on the two-state model using the following parameters; K_R (basal R/R^*) = 0.02; for ICI-118,551, $K_L = 5 \times 10^{-10}$ M, and $K_L^* = 3 \times 10^{-6}$ M; pindobind was assumed to alter K_R (basal R/R^*) from 0.02 to 0.6 as a result of preferential enrichment of R. Virtually identical simulations were obtained including receptor reserve using the same ligand parameters and $K_R = 10$ and $K_E = 0.002$. In this case, pindobind was assumed to reduce the absolute value of R and R^* equally by 30-fold as a result of receptor alkylation. Experimentally, ICI-118,551 produced a maximum decrease in tension of 80%, and for the simulations, it was assumed that this value represented full inverse agonism. To allow side-by-side comparison of the theoretical and experimental graphs, the y-axes of the simulations were scaled to this range.

70% after 4 min (Fig. 4a) with more profound decreases after longer treatment periods (data not shown). This was associated with a fall in heart rate (485 ± 13 to 336 ± 17 b.p.m.) and a fall in LV systolic pressure (74 ± 2 to 54 ± 6 mm Hg) (data not shown). ICI-118,551 had no effect on LV $dP/dt_{\max}$ in control animals (data not shown). Alprenolol produced only a 25% decrease in LV $dP/dt_{\max}$ (Fig. 4a) with no further effect after 4 min, and no significant effects on either heart rate or LV systolic pressure. In addition, pre-treatment with alprenolol significantly blocked the effects of ICI-118,551 on LV $dP/dt_{\max}$ (Fig. 4a) as well as heart rate and LV systolic pressure.

The effect of ICI-118,551 and alprenolol on adenylyl cyclase activity was also examined in cardiac membranes²⁴. β -Adrenoceptor stimulation increases adenylyl cyclase activity, increasing intracellular cAMP and contractility; baseline levels of adenylyl cyclase were elevated in transgenic animals²⁴. Comparable to the results obtained in isolated atria and *in vivo*, alprenolol could antagonize the inhibitory effect of ICI-118,551 on adenylyl cyclase activity (Fig. 4b); co-incubation of both antagonists resulted in a significant attenuation of the effect of ICI-118,551 alone.

These experiments suggest that the myocardium of transgenic mice contains a markedly elevated number of wild-type β_2 -adrenoceptors in the active conformation (R^*), which can therefore increase adenylyl cyclase activity and enhance left atrial tension and left ventricular contractility. These findings provide physiological support for a model of receptor action which predicts that receptors exist in an equilibrium between two allosterically different conformations (Fig. 1), R and R^* (refs 5, 11–17). In addition, these data demonstrate that ICI-118,551

functions as an inverse agonist, presumably, by shifting the equilibrium between R and R^* towards the inactive conformation (R).

Using equations based on a simple two-state model of receptor activation (Fig. 1), the interaction between the effects of the inverse agonist, ICI-118,551, and the agonist isoprenaline (Fig. 2c) or the competitive (neutral) antagonist alprenolol (Fig. 3b) could be quantitatively modelled. The striking correlation between the simulations and the experimental data demonstrates that this model accurately predicts the behaviour of ligands in this system. In accepting the theoretical analysis of the data, the choice of parameter values used is important and it should be emphasized that the values of K_L and K_L^* for ICI-118,551, and the value of the basal ratio of R/R^* (K_R) were the same in each of the three simulations.

Classical receptor theory has as its cornerstone the occupancy–response relationship^{1,3}, that is, the magnitude of the response obtained is directly proportional to the number of receptors occupied by ligand. The increased potency of the inverse agonist response after receptor alkylation by pindobind, however, is irreconcilable with a single-state receptor model for occupancy–response. The simulations using the two-state model accurately predict the effects of pindobind on ICI-118,551 response curves.

The prevailing concept of ligand efficacy for G-protein-coupled receptors may now need to be carefully re-examined. With the development of sensitive test systems for the detection of inverse agonism, such as these transgenic animals, many pharmacological agents may need reclassification. If ligand efficacy is the differential affinity of a ligand for the two existing conformations of the receptor, then an expectation of 'zero' efficacy, which would be required of a pure neutral antagonist, may be relatively uncommon, because this would require identical drug affinities for both receptor states^{11,12}. Numerous drugs previously classified as neutral antagonists, therefore, may actually function as inverse agonists^{11,12,15,16,21,22}, and such a reclassification would have important pharmacological and therapeutic implications.

These studies also highlight the important potential for inverse agonists in drug development as new, targeted therapeutic agents. In addition to eliminating tonic endogenous agonist tone like the neutral antagonists, inverse agonists can also negate constitutive receptor activity; this may be particularly important in disease states that result from constitutively activating receptor mutations. The development of thyroid adenomas and male precocious puberty^{27,28} are two examples of such human diseases in which receptor mutants play a role in disease pathogenesis by inducing constitutive activity. Potentially these compounds would also have applicability in diseases resulting from receptor overexpression, such as the dopamine D_4 receptor in schizophrenia²⁹; such overexpression can induce a constitutively active phenotype²⁴ and make the disease resistant to standard antagonists. Inverse agonists, therefore, may represent an important and specific therapeutic approach for such disease states. □

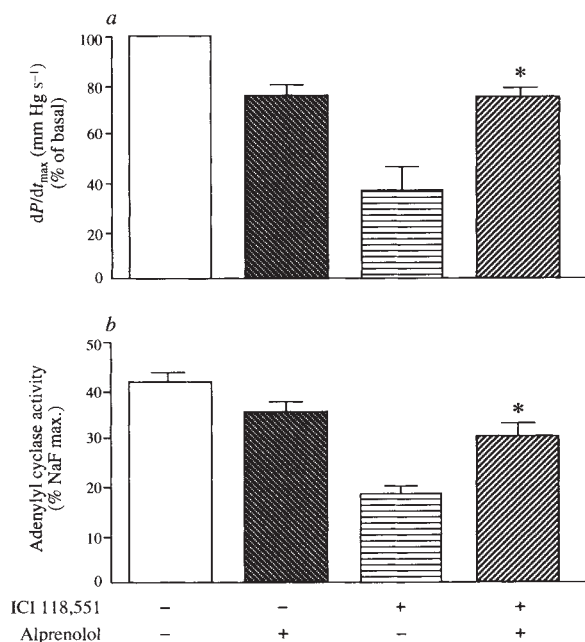


FIG. 4 a, The effect of ICI-118,551 and alprenolol on the *in vivo* measurement of LV $dP/dt_{\max}$ ²⁴ in TG-4 mice. LV $dP/dt_{\max}$ in TG-4 mice ($n=7$) before and 4 min after the administration of ICI-118,551 ($5 \mu\text{g i.v.}$) or alprenolol ($10 \mu\text{g i.v.}$); at 6 min (data not shown), all animals treated with ICI-118,551 became hypotensive and developed severely depressed LV $dP/dt_{\max}$. The effect of ICI-118,551 on LV $dP/dt_{\max}$ in TG-4 animals pretreated with alprenolol 4 min before the administration of ICI-118,551 ($n=8$); *, the inhibitory effect of ICI-118,551 was significantly attenuated by alprenolol pretreatment, $P<0.05$. Data are reported as the mean $\pm$ s.e.m. b, The effect of ICI-118,551 (1×10^{-7} M) and alprenolol (1×10^{-6} M) on adenylyl cyclase activity²⁴ in membranes prepared from TG-4 hearts; *, differs from response in the presence of ICI-118,551 alone, $P<0.05$. All responses are expressed as a percentage of the maximal stimulation by NaF (10 mM) ($n=4$ or 5 per group).

Received 3 June 1994; accepted 16 January 1995.

- Clark, A. J. in *The Mode of Action of Drugs on Cells* (Edward Arnold, London, 1933).
- Ahrens, E. J. *Arch. Int. Pharmacodyn.* **99**, 32–49 (1954).
- Stephenson, R. P. *Br. J. Pharmacol.* **11**, 379–393 (1956).
- De Lean, A., Stadel, J. M. & Lefkowitz, R. J. *J. Biol. Chem.* **255**, 7108–7117 (1980).
- Samama, P., Cotecchia, S., Costa, T. & Lefkowitz, R. J. *J. Biol. Chem.* **268**, 4625–4636 (1993).
- Ehlert, F. J. *Trends Pharmacol. Sci.* **7**, 28–32 (1986).
- Costa, T., Ogino, Y., Munson, P. J., Onaran, O. & Rodbard, D. *Molec. Pharmacol.* **41**, 549–560 (1992).
- Onaran, H. O., Costa, T. & Rodbard, D. *Molec. Pharmacol.* **43**, 245–255 (1992).
- Costa, T. & Herz, A. *J. Proc. natn. Acad. Sci. U.S.A.* **86**, 7321–7325 (1989).
- Costa, T., Lang, J., Gless, C. & Herz, A. *Molec. Pharmacol.* **37**, 383–394 (1990).
- Samama, P., Pei, G., Costa, T., Cotecchia, S. & Lefkowitz, R. J. *Molec. Pharmacol.* **45**, 390–394 (1994).
- Chidiac, P., Hebert, T. E., Valiquette, M., Dennis, M. & Bouvier, M. *Molec. Pharmacol.* **45**, 490–499 (1994).
- Lefkowitz, R. J., Cotecchia, S., Samama, P. & Costa, T. *Trends Pharmacol. Sci.* **14**, 303–307 (1993).
- Schutz, W. & Freissmuth, M. *Trends Pharmacol. Sci.* **13**, 376–380 (1992).

15. Mewes, T., Dutz, S., Ravens, U. & Jakobs, K. H. *Circulation* **88**, 2916–2922 (1993).
16. Gotze, K. & Jakobs, K. H. *Eur. J. Pharmacol.* **268**, 151–158 (1994).
17. Monod, J., Wyman, J. & Changeux, J.-P. *J. molec. Chem.* **12**, 88–118 (1965).
18. Colquhoun, D. in *Drug Receptors* 149–181 (MacMillan, London, 1973).
19. Leff, P. *Trends pharmacol. Sci.* (in the press).
20. Cerione, R. A. et al. *Biochemistry* **23**, 4519–4525 (1984).
21. Senogles, S. E. et al. *J. biol. Chem.* **262**, 4860–4867 (1987).
22. Barker, E. L., Westphal, R. S., Schmidt, D. & Sanders-Bush, E. J. *J. biol. Chem.* **269**, 11687–11690 (1994).
23. Tian, W.-N., Duzic, E., Lanier, S. M. & Deth, R. C. *Molec. Pharmacol.* **45**, 524–531 (1994).
24. Milano, C. A. et al. *Science* **264**, 582–586 (1994).
25. Arch, J. R. S. & Kaumann, A. J. *Mednl Res. Rev.* **13**, 663–729 (1993).
26. Schwartz, D. D. & Eikenburg, D. C. *J. Pharmacol. exp. Ther.* **244**, 11–18 (1988).
27. Shenker, A. et al. *Nature* **365**, 652–654 (1993).
28. Parma, J. et al. *Nature* **365**, 649–651 (1993).
29. Seeman, P., Guan, H.-C. & Van Tol, H. H. M. *Nature* **365**, 441–445 (1993).

ACKNOWLEDGEMENTS. We thank D. Eikenburg for his support and critical review of this manuscript and M. Holben and D. Addison for secretarial assistance.

Replication of transcriptionally active chromatin

Renzo Lucchini & José M. Sogo

Institute of Cell Biology, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

In eukaryotic cells, active genes and their regulatory sequences are organized into open chromatin conformations in which nucleosomes can be modified, disrupted or totally absent^{1–3}. It has been proposed that these characteristic chromatin structures and their associated factors might be directly inherited by the newly synthesized daughter strands during chromosome duplication^{4–6}. Here we show that in the yeast *Saccharomyces cerevisiae*, replication machinery entering upstream of a transcriptionally active ribosomal RNA gene generates two newly replicated coding regions regularly packaged into nucleosomes, indicating that the active chromatin structure cannot be directly inherited at the replication fork. Whereas the establishment of an exposed chromatin conformation at some newly replicated rRNA gene promoters can occur shortly after the passage of the replication fork, regeneration of the active chromatin structure along the coding region is always a post-replicative process involving disruption of preformed nucleosomes.

We have previously shown that in a growing yeast cell only a fraction of the ~150 tandemly repeated rRNA genes are transcriptionally active⁷. Active and inactive copies can be distinguished by their different accessibility *in vivo* to the intercalating drug psoralen, which can introduce crosslinks into DNA sites that are not protected by nucleosomes^{3,8}. In this assay, each rRNA coding fragment is separated into two distinct bands on native agarose gels: a slowly migrating band representing highly crosslinked DNA derived from transcriptionally active genes organized in a heavily disrupted chromatin structure, and a fast-migrating band containing slightly crosslinked DNA originating from the inactive, nucleosome-packed gene copies^{9,10} (s and f bands in Fig. 1, lanes L).

To study the chromatin structure of newly replicated DNA, we analysed the extent of crosslinking of rRNA coding fragments excised from the replicated branches of replication intermediates purified from preparative two-dimensional gels¹¹. Figure 1b (lane 3) shows an analysis of the 1.9-kilobase (kb) *EcoRI*-digested coding fragment derived from the replicated sections of gel-purified replication intermediates (shaded portions of fork 3 in Fig. 1a), in which there is only one band with the same mobility as the slightly crosslinked, fast-migrating band (for exponentially growing cells, see Fig. 1d). On the other hand, the proportion of the s and f bands generated by a fragment derived from the unreplicated portion of early replication intermediates eluted from the same two-dimensional gel is consistent with the presence of a large fraction of replicating molecules

having parental DNA sequences organized in an active chromatin structure (Fig. 1c, lane 1). These results indicate that the replication machinery always generates two newly replicated coding regions tightly packaged in nucleosomal arrays.

To examine chromatin structure at the replication forks in more detail, we analysed single DNA replication intermediates by electron microscopy^{3,12}. The nucleosomal intergenic spacer is visualized as a row of single-stranded bubbles, whereas highly crosslinked DNA derived from transcriptionally active genes has a double-stranded appearance⁷ (Fig. 2a). In the rDNA repeats shown in Fig. 2b and c, a replication fork has entered the upstream gene. Whereas the unreplicated parental DNA immediately ahead of the fork is fairly continuously crosslinked, the newly replicated coding sequences are organized as rows of single-stranded bubbles, consistent with the presence of nucleosomes (see also Fig. 2d). Measurements of the length of the short nucleosome-free region immediately behind the replication

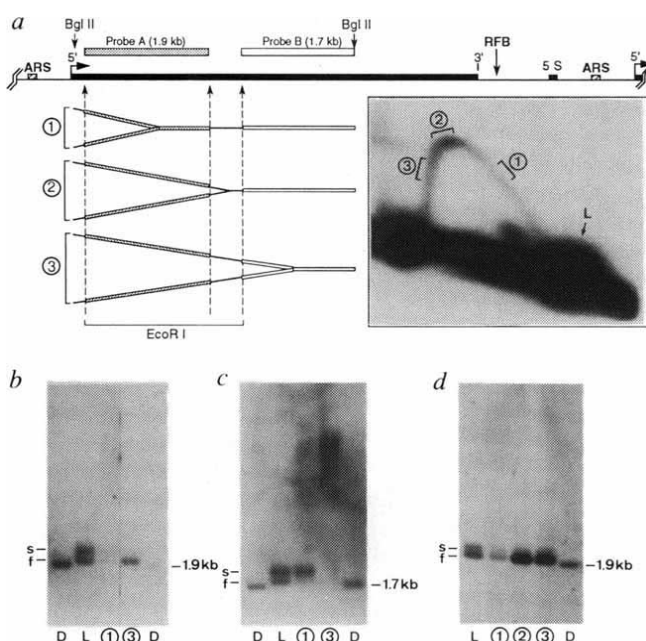


FIG. 1 Psoralen accessibility of replicating rRNA coding fragments. a, Map showing the structural organization of the rDNA repeat unit of *S. cerevisiae*. The 35S precursor coding region is indicated as a filled box. In the rRNA intergenic spacer, the replication fork barrier (RFB), the 5S rRNA gene (5S) and the autonomously replicating sequence (ARS) are shown. Probes A and B detect the 1.9-kb *EcoRI* and the 1.7-kb *EcoRI*-*BglIII* fragments, respectively, derived from the tandemly repeated rRNA coding regions. The autoradiograph below the map shows rDNA from psoralen crosslinked S-phase cells which has been digested with *BglIII*, separated on a 2D gel, blotted and hybridized with probe A. The position of the linear 4.5-kb *BglIII* coding fragment is indicated by the arrow (L). As in yeast most of the rRNA genes are replicated in the same direction as transcription¹⁴, most of the Y-shaped replication intermediates of this coding fragment migrating in regions 1, 2 and 3 of this arc have the same structure and orientation as those of forks 1, 2 and 3 (shown on the left), respectively. b, rDNA isolated from regions L, 1 and 3 of a 2D gel similar to that shown in a, was digested with *EcoRI*, separated on a 1.5% agarose gel, blotted and hybridized with probe A. Lane D: non-crosslinked control DNA digested with *EcoRI*. c, Aliquots of the same samples as used in b were separated on the same gel, blotted and hybridized with probe B. d, Similar experiment as in b, except that the rDNA was isolated from psoralen-crosslinked exponentially growing cells.

METHODS. *S. cerevisiae* A1 cells (MATa ade2-101 ura3-52 his3Δ200 lys2-801 Δbar1::LYS2) were grown, synchronized, and photoreacted as described¹¹. Crosslinked rDNA was purified on a CsCl-actinomycin D gradient⁷. Preparative 2D-gel electrophoresis, transfer and hybridization were done as described¹¹.

fork¹² gave similar values for replication forks entering nucleosomal intergenic spacers (288 ± 139 base pairs (bp) or transcriptionally active coding regions (257 ± 149 bp; data from 12 and 27 forks, respectively). These data indicate that the rapidity with which nucleosomal structures are assembled on the newly

replicated DNA is always the same, independently of whether replication machinery is advancing through an active or an inactive chromatin template.

When the rightwards moving fork of replication intermediates of the type shown in Fig. 2d passes the *PvuI* site, these bubbled

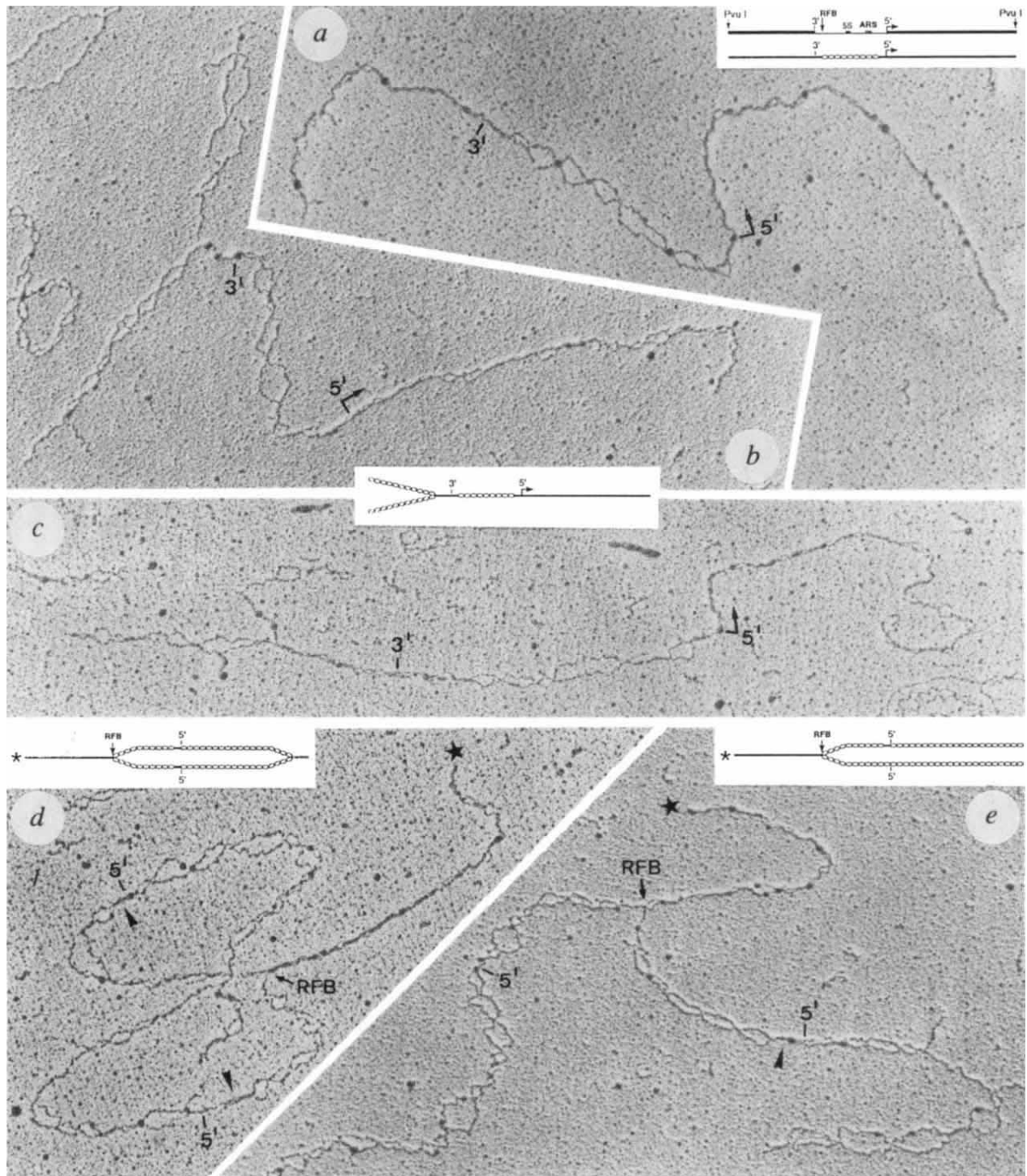


FIG. 2 Chromatin structure at replication forks moving into active rRNA coding regions. rDNA purified from psoralen crosslinked S-phase cells was digested with *PvuI*, which cuts only once per repeating unit (inset in a), and spread for electron microscopy under denaturing conditions³. a, Electron micrograph of a non-replicating 9.1-kb rDNA repeat. The putative 3' and 5' ends of the two adjacent rRNA genes are indicated. b and c, Micrographs showing rDNA repeats being replicated by a fork

that has entered from upstream. d, In this repeat, the origin of replication located in the rRNA intergenic spacer has fired: the leftwards-moving fork has stopped at the replication fork barrier (RFB)^{13,14}, whereas the rightwards-moving fork is replicating the downstream gene. e, Y-shaped replication intermediate containing a fork arrested at the RFB. Arrowheads in d and e point to short, heavily crosslinked DNA stretches close to the putative 5' ends of the newly replicated coding regions.

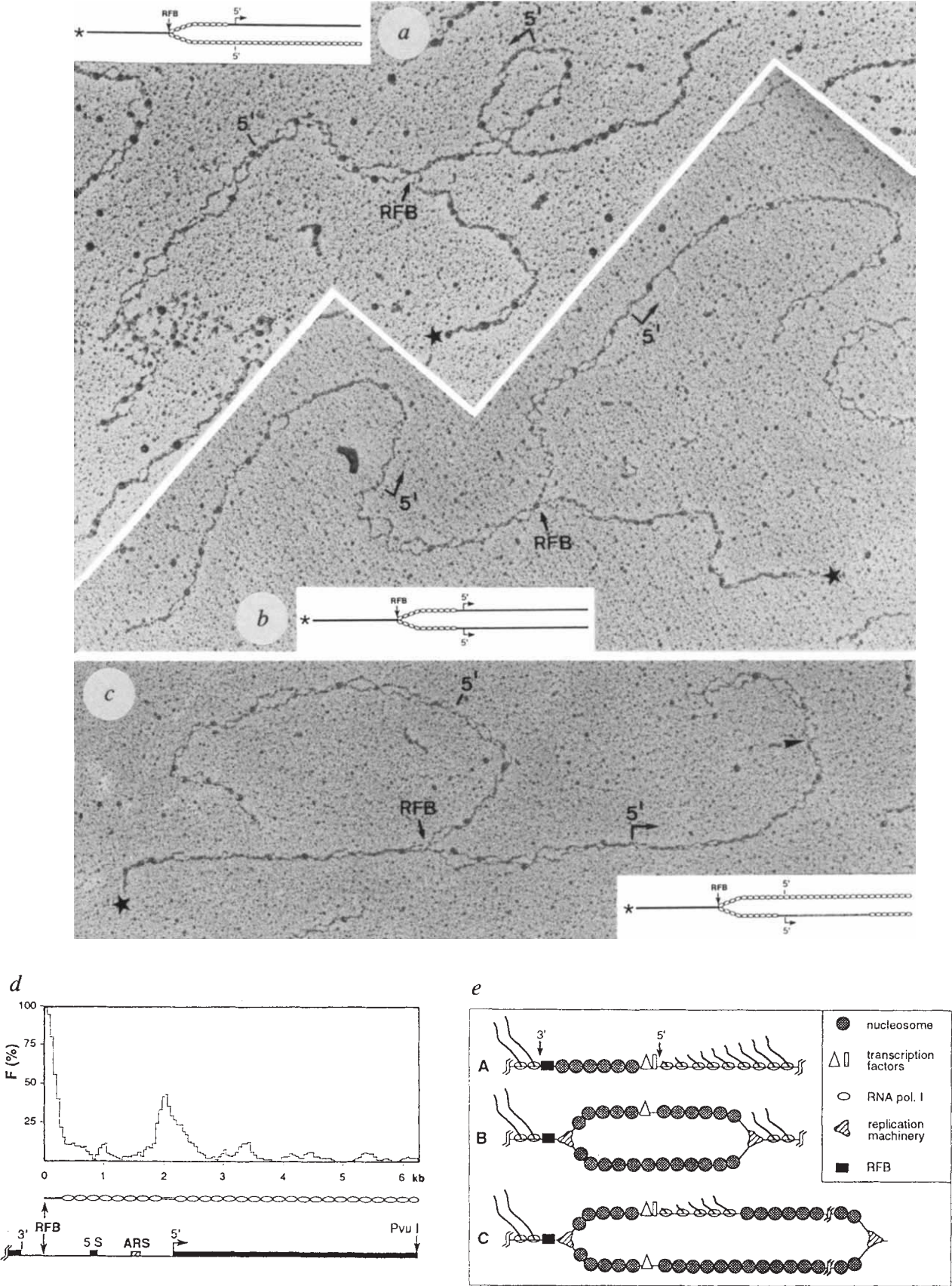


FIG. 3 Reactivation of replicated rRNA genes. *PvuI* digested rDNA from psoralen crosslinked S-phase cells was prepared as in Fig. 2. The three micrographs represent Y-shaped replication intermediates containing a replication fork arrested at the RFB. In the molecules shown in *a* and *b*, one or both replicated genes, respectively, appear to be fairly continuously crosslinked, indicative of active rRNA transcription. *c*, Arrested replication intermediate, in which one of the replicated rRNA coding region is organized partly as heavily crosslinked, double-stranded DNA and partly in single-stranded bubbles. The transition point between these two distinct regions is indicated by the arrowhead. *d*, Mapping of the short, heavily crosslinked DNA stretches on the replicated rDNA. Heavily crosslinked, double-stranded DNA stretches larger than 150 bp were mapped on the replicated portions of replication intermediates of the type shown in Fig. 2*d* and *e* and in *a*. The analysis was restricted to replicated DNA of 93 molecules in which most of the coding region was organized in single-stranded bubbles. The distance from the branch point of the arrested fork (RFB) and the right end of the molecule (*PvuI* site) was normalized to 6.25 kb (see map below the histogram). The value *F* on the ordinate indicates the percentage of the molecules that had a double-stranded DNA stretch larger than 150 bp in a given length increment of 50 bp. The short stretch of non-nucleosomal DNA at the arrested replication fork was also included in the analysis. *e*, Model for rRNA gene inactivation and reactivation during chromatin replication. The three drawings show one series of events that might happen during duplication of an active rRNA gene. (1) Portions of two adjacent transcriptionally active genes separated by a nucleosomal intergenic spacer. (2) The origin of replication in the spacer has fired: the leftwards moving fork has arrested at the RFB, whereas the rightwards-moving replication machinery has entered the downstream gene. Both newly replicated coding regions behind the fork are always rapidly packaged in nucleosomes. On one of the two newly replicated daughter strands, a transcription factor is bound at the gene promoter, keeping the chromatin structure in an open conformation ('potentially active promoter'). At this stage, we also found molecules in which both newly replicated promoters were either nucleosome free (Fig. 2*d*), or nucleosome packed (data not shown). (3) After duplication of the parental gene, a functional transcription initiation complex has been established at the previously open promoter, and transcription initiation has occurred. The wave of the newly initiated RNA polymerases open up the chromatin structure along the transcription unit. Meanwhile, a transcription factor has bound to the sibling strand, disrupting the chromatin structure at the gene promoter. Finally, transcription initiation on the sibling promoter might also occur.

replicative structures are converted to long lived, Y-shaped molecules, as the replication fork directed to the left is arrested at the replication fork barrier^{13,14} (Fig. 2*e*). Whereas in the majority of these molecules, most of the duplicated DNA was organized in single-stranded bubbles (Table 1), in some replication intermediates one or both replicated sibling genes had a double-stranded appearance indistinguishable from that of transcriptionally active genes (Fig. 3*a* and 3*b*). Furthermore,

from 201 molecules analysed, seven replicative structures had one of the replicated gene organized into two distinct domains of variable sizes: an upstream portion that was fairly continuously crosslinked and a downstream portion organized in single-stranded bubbles (Fig. 3*c*). From this result, we propose that regeneration of the active chromatin structure is a processing event mediated by the newly initiated RNA polymerase I molecules advancing through the template and disrupting nucleosomes in their path.

A closer inspection of newly replicated rDNA revealed the presence of short, well defined stretches of heavily crosslinked DNA located at the position of the rRNA gene promoter (arrowheads in Fig. 2*d* and *e*; see also Fig. 3*d*). As these open promoters are followed by a coding region packaged in an inactive chromatin structure, we classify them as potentially active. The fact that these open promoters are already apparent shortly after the passage of the replication fork (Fig. 2*d*) raises the possibility that these potentially active chromatin structures are assembled at the instant of replication. It has been proposed that the local perturbation of parental chromatin at the replication fork^{12,15}, followed by the short temporal lag in nucleosome reformation¹², might offer a unique opportunity for transcription factors to bind to nascent DNA before it is sequestered into inaccessible nucleosomal structures^{16,17}. If the instant of replication is the only period during which the cell can establish these transcriptionally competent structures, one should expect to find the same percentage of open promoters on the replicated sections of early (bubbles) and late (arrested Y-shaped) replication intermediates. As the frequency of exposed rRNA gene promoters (active and potentially active) on the arrested replication intermediates is much higher than the one calculated for bubbled replicative structures (Table 1), we conclude that a considerable fraction of these exposed chromatin conformations is established post-replicatively by a mechanism involving disruption of preformed nucleosomes. The staggered appearance of open promoters and coding regions following the passage of the replication fork suggests a two-step activation process in which a local disruption of chromatin over the rRNA gene promoter precedes the assembly of a functional transcription initiation complex (Fig. 3*e*). Disruption of chromatin at the gene promoter might be mediated by the REB1/GRF2 factor, which has a strong binding site located ~200 bp upstream of the transcription initiation site^{18,20}.

Our main conclusion is that in this particular gene system, dramatic changes in the basic repeating structure of chromatin related to the transcriptional process do not represent structural information that can be directly transmitted to the sister chromatids during chromosome duplication. Replicating chromatin sections seem rather to undergo a 'structural resetting' whereby

TABLE 1 Summary of rDNA replication intermediates analysed in the electron microscope

Type of replication intermediates	No. of molecules analysed	No. of replicated promoters*			Replicated active + potentially active promoters
		Active	Potentially active	Inactive	
Bubbles (Fig. 2 <i>c</i>)	20†	0	8	32	20%
Arrested Y-shaped					
Both replicated genes inactive (Fig. 2 <i>d</i>)	57	38	34	104	41%
Only one replicated gene active (Fig. 3 <i>a</i>)	24‡				
Both replicated genes active (Fig. 3 <i>b</i>)	7				
Y-shaped with rightwards moving fork (Fig. 2 <i>b</i>)	25†	—	—	—	—

Data derive from electron microscopy analysis of an aliquot of the rDNA sample examined in Fig. 1*b* and *c*.

* Active promoters are those flanked by heavily crosslinked coding regions. Potentially active promoters are organized in an open chromatin conformation and are flanked by nucleosomal coding regions. Their number was estimated from the percentage of replicated, inactive gene portions that had a stretch of heavily crosslinked DNA at their 5' ends (from histogram like that in Fig. 3*d*). The number of inactive promoters was obtained by subtracting the number of potentially active promoters from the total number of replicated coding regions organized in single-stranded bubbles.

† Among type-1 and type-3 molecules, only two examples showed a replication fork entering a nucleosomal coding region, whereas, in all of the molecules analysed, the newly replicated DNA behind the fork was packaged in nucleosomes.

‡ In three molecules the gene had been recently activated (Fig. 3*c*).

nascent DNA appears to be always organized into regular nucleosomal arrays characteristic of the transcriptionally inactive state. However, our results do not exclude the possibility that some kind of epigenetic information escaping our analysis might be transmitted at the instant of replication. □

Received 7 November 1994; accepted 3 February 1995.

1. Felsenfeld, G. *Nature* **355**, 219–224 (1992).
2. Kornberg, R. D. & Lorch, Y. A. *Rev. Cell Biol.* **8**, 563–587 (1992).
3. Sogo, J. M., Ness, P. J., Widmer, R. M., Parish, R. W. & Koller, T. *J. molec. Biol.* **178**, 897–928 (1984).
4. Alberts, B., Worcel, A. & Weintraub, H. *Proc. int. Symp. on the Eukaryotic Genome* (eds Bradbury, M. & Javaherian, K.) 165–191 (Academic, New York, 1977).
5. Brown, D. D. *Cell* **37**, 359–365 (1984).
6. Weintraub, H. *Cell* **42**, 705–711 (1985).
7. Dammann, R., Lucchini, R., Koller, T. & Sogo, J. M. *Nucleic Acids Res.* **10**, 2331–2338 (1993).
8. Conconi, A., Losa, R., Koller, T. & Sogo, J. M. *J. molec. Biol.* **178**, 920–928 (1984).
9. Conconi, A., Widmer, R. M., Koller, T. & Sogo, J. M. *Cell* **57**, 753–761 (1989).
10. Lucchini, R. & Sogo, J. M. *Molec. cell. Biol.* **12**, 4288–4296 (1992).
11. Lucchini, R. & Sogo, J. M. *Molec. cell. Biol.* **14**, 318–326 (1994).
12. Sogo, J. M., Stahl, H., Koller, T. & Knippers, R. *J. molec. Biol.* **189**, 189–204 (1986).
13. Brewer, B. J. & Fangman, W. L. *Cell* **55**, 637–643 (1988).
14. Linskens, M. H. K. & Huberman, J. A. *Molec. cell. Biol.* **8**, 4927–4935 (1988).
15. Jackson, V. *Biochemistry* **29**, 719–731 (1990).
16. Svaren, J. & Chalkley, R. *Trends Genet.* **6**, 52–56 (1990).
17. Wolffe, A. P. *J. Cell Sci.* **99**, 201–206 (1991).
18. Kulkens, T., van Heerikhuizen, H., Klootwijk, J., Oliemans, J. & Planta, R. J. *Curr. Genet.* **16**, 351–359 (1989).
19. Morrow, B. E., Johnson, S. P. & Warner, J. R. *J. biol. Chem.* **264**, 9061–9068 (1989).
20. Chasman, D. I. *et al. Genes Dev.* **4**, 403–514 (1990).

ACKNOWLEDGEMENTS. We thank M. Aebi for the yeast cell line A1 and for help with cell synchronization; J. Szostak and R. Wu for plasmids pSZ26 (from which probes A and B were obtained); and H. Mayer-Rosa for technical assistance. This work was supported by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung.

Association of Cdk-activating kinase subunits with transcription factor TFIIF

Hiroaki Serizawa, Tomi P. Mäkelä,
Joan Wellky Conaway*, Ronald C. Conaway*,
Robert A. Weinberg & Richard A. Young

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142 and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* Program in Molecular and Cell Biology, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104, USA

THE RNA polymerase II large subunit contains an essential carboxy-terminal domain (CTD) believed to be involved in the response to regulators during transcription initiation^{1–10}. The CTD is phosphorylated on a portion of RNA polymerase II molecules *in vivo*^{11,12} and it can be phosphorylated by the general transcription factor TFIIF *in vitro*^{13–15}. A highly purified TFIIF from rat liver has been described¹⁶; this, like human and yeast TFIIF, contains associated CTD kinase and helicase activities^{13–18}. We report here that two polypeptides of the purified mammalian TFIIF are the MO15/Cdk7 kinase and cyclin H subunits of the Cdk-activating kinase Cak^{19–21}, previously identified as a positive regulator of Cdc2 and Cdk2. TFIIF and Cak preparations are each capable of phosphorylating recombinant CTD and recombinant Cdk2 proteins. The presence of Cak in TFIIF indicates that Cak may have roles in transcriptional regulation and in cell-cycle control.

We investigated the possibility that mammalian TFIIF is associated with a Cdk-related kinase because we recently learned that *Saccharomyces cerevisiae* TFIIF is associated with such a kinase (R. D. Kornberg, personal communication). Highly

purified TFIIF from rat liver was found to contain a kinase capable of phosphorylating recombinant Cdk2, a behaviour ascribed to the Cdk-activating kinase Cak¹⁹. To obtain additional evidence for the association of Cak with TFIIF, fractions from the final column of the purification were analysed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis

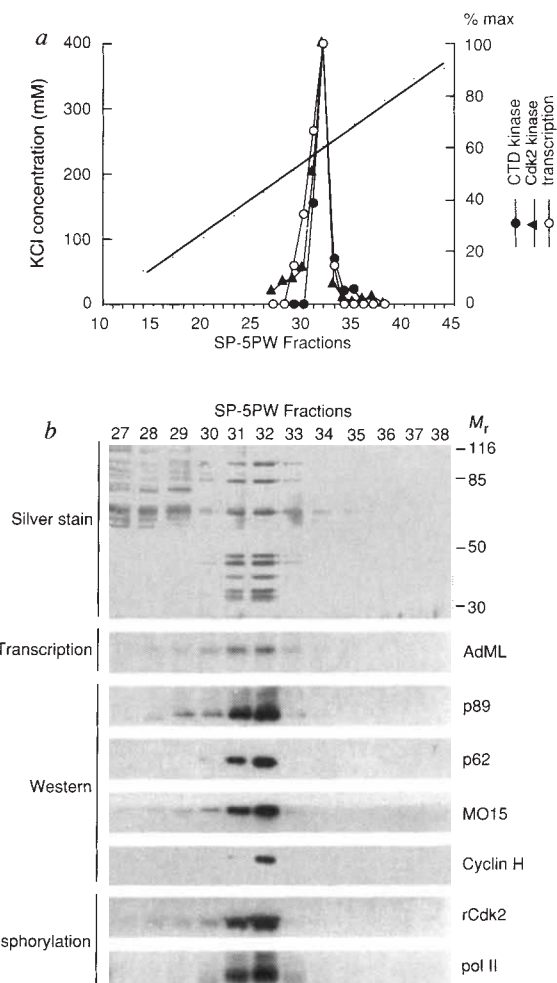
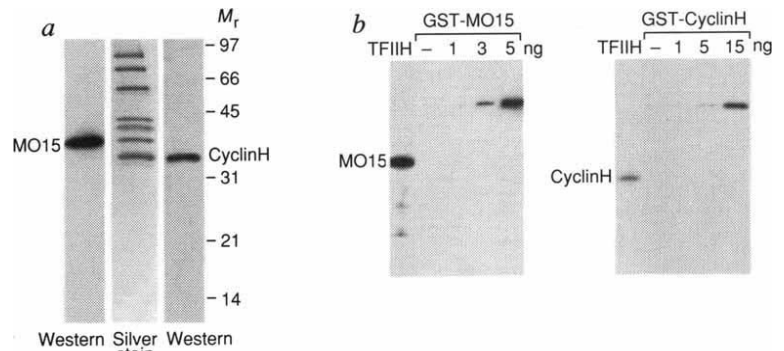


FIG. 1 Copurification of MO15 and cyclin H with TFIIF. *a*, Coelution of CTD kinase, Cdk2 kinase and TFIIF transcription activities from a SP-5PW column. *b*, Analysis of SP-5PW fractions by SDS-PAGE (silver stain), in TFIIF-dependent transcription from the adenovirus major late promoter (Transcription), by western blot with antibodies against the p89 and p62 subunits of TFIIF and with antibodies against MO15 and cyclin H, and in phosphorylation assays with recombinant Cdk2 and RNA polymerase II. Experiments reported previously show that the CTD of RNA polymerase II is phosphorylated by the rat liver TFIIF preparation¹⁴. METHODS. Rat liver TFIIF was purified as described¹⁷. The transcription assay (30 µl) was as described^{27,28} with a template containing the adenovirus major late promoter (AdML), and runoff transcripts were analysed on 6% polyacrylamide/7M Urea gels. Western blot experiments were as described²⁹. Anti-MO15 and anti-cyclin H rabbit polyclonal antibodies were raised against 15-amino-acid C-terminal peptides coupled to keyhole limpet haemocyanin (KLH). These antibodies do not react with other known kinases or cyclins. Anti-TFIIF p89 (ERCC3) and p62 antibodies were from Santa Cruz Biotech. Bands on the western blots were visualized by chemiluminescence with horseradish peroxidase-conjugated secondary antibody (Amersham). CTD and Cdk2 kinase assays were as described¹⁴ with rat liver RNA polymerase II (0.001 units) and 5 µg of GST-Cdk2²¹. Transcriptional and kinase activities were quantified by densitometry and with a Fijix Bio-image Analyzer. Portions of this and subsequent figures were prepared from digital replicas of primary data scanned using a UMAX UC840 Max Vision digital scanner.

FIG. 2 Identification and stoichiometry of MO15 and cyclin H components of rat liver TFIH. *a*, Identification of MO15 and cyclin H polypeptides among the subunits of TFIH by silver stain and western blot analysis. Three identical aliquots of ~100 ng of TFIH were subjected to SDS-PAGE. One lane was silver stained and the other two lanes were transferred to PVDF membranes and probed with anti-MO15 and anti-cyclin H antibodies as described in Fig. 1. The MO15 antibody reacts with p40 and the cyclin H antibody reacts with p34. *b*, Quantification of the levels of MO15 and cyclin H polypeptides in rat liver TFIH by western blot analysis. About 100 ng of TFIH (containing ~10 ng of p40 and 10 ng of p34) was subjected to western blot analysis together with various amounts of recombinant GST-MO15 or GST-cyclin H. The levels of TFIH polypeptides were estimated on the basis of both the A_{280} of the preparation and on Coomassie brilliant blue staining of the subunits. The signals obtained with anti-MO15 and anti-cyclin H antibodies suggest that much, if not all, of the p40 and p34 polypeptides in the TFIH preparation are MO15 and cyclin H.



(SDS-PAGE), were subjected to western blot analysis with antibodies against MO15 and cyclin H, and were used in kinase assays using CTD and Cdk2 substrates (Fig. 1). The SDS-PAGE analysis revealed that the rat liver TFIH preparation contains at least eight major polypeptides that cofractionate with transcription activity. Western blot analysis indicated that polypeptides recognized by MO15 and cyclin H antibodies copurified with the p89 (ERCC3) and p62 subunits of TFIH. The kinase assays revealed that RNA polymerase II CTD and Cdk2 phosphorylation activities coeluted with these polypeptides and the transcriptional activity of the TFIH preparation.

Further analysis of the rat liver TFIH preparation revealed that the MO15 and cyclin H antibodies recognize p40 and p34, respectively (Fig. 2*a*). Comparison of the western blot signals obtained with known amounts of recombinant MO15 and cyclin H together with known amounts of rat liver TFIH indicated that most, if not all, of the p40 and p34 polypeptides are MO15 and cyclin H, respectively (Fig. 2*b*). These results suggest that MO15 and cyclin H may be stoichiometric components of TFIH. The MO15 and cyclin H antibodies used in this analysis can be blocked by their respective recombinant antigens and do not crossreact with any of the known Cdks or cyclins. In addition, antibodies that react with other known Cdks do not produce signals with the TFIH preparation. Time-of-flight mass spectrometry revealed that peptides obtained by trypsin digestion of p40 and p34 are those predicted for MO15 and cyclin H, confirming that the p40 and p34 polypeptides of rat liver TFIH contain MO15 and cyclin H sequences (data not shown).

If MO15 and cyclin H were subunits of TFIH, then TFIH subunits and transcriptional activity should co-immunoprecipitate with MO15. Indeed, TFIH transcriptional activity and the p89 (ERCC3) subunit of TFIH was found to be immunoprecipitated with anti-MO15 antibody (Fig. 3). Cdk2 kinase activity was also found to coprecipitate with these polypeptides. These experiments confirm that the Cdk-activating kinase Cak is a component of rat liver TFIH.

If TFIH contains functional Cak, then it should be able to activate Cdk2, enabling Cdk2 to phosphorylate the test substrate histone H1. Figure 4*a* shows that the rat liver TFIH preparation was indeed able to activate a glutathione-S-transferase (GST)-Cdk2/cyclin A preparation to phosphorylate histone H1. If the Cak associated with TFIH is functionally equivalent to that regulating the cell-cycle clock, then Cak preparations isolated by virtue of their ability to phosphorylate CDK substrates should be capable of phosphorylating the CTD. Most of the Cak activity in cell lysates occurs in a complex of about 175K²⁰, considerably smaller than TFIH. Immunoprecipitated Cak preparations from human cells can phosphorylate recombinant CTD and recombinant Cdk2 (Fig. 4*b*). In a control experiment,

both CTD and Cdk2 phosphorylation activities were lost when the Cak immunoprecipitation was done in the presence of an excess of the synthetic MO15 peptide used to generate the anti-MO15 antibody.

Activation of Cdk2 by Cak involves phosphorylation of the threonine 160 of Cdk2 (ref. 22). In contrast, serine residues have been reported to be phosphorylated upon incubation of CTD with TFIH¹⁷. If Cak is the kinase incorporated into TFIH, then TFIH and Cak preparations might both be expected to phosphorylate threonine on Cdk2, yet phosphorylate serine in the CTD. The data in Fig. 4*c* show that this is the case.

The RNA polymerase II CTD has been implicated in the response of the transcription apparatus to regulatory signals *in vivo*⁴⁻⁸. Both *in vivo* and *in vitro* experiments suggest that the CTD is generally found in an unphosphorylated form in RNA polymerase II complexes that are poised to initiate transcription whereas the CTD of RNA polymerase II that is in the midst of transcriptional elongation is often highly phosphorylated^{3,11,23}. Thus, CTD phosphorylation has been proposed to be a regulatory event governing some aspect of the transcription initiation process. Biochemical studies suggest that CTD phosphorylation can be mediated by the general transcription factor TFIH^{13,15}, which has led to the notion that TFIH acts to control RNA polymerase II function in part through its ability to cause CTD phosphorylation. This work reveals that two of the protein com-

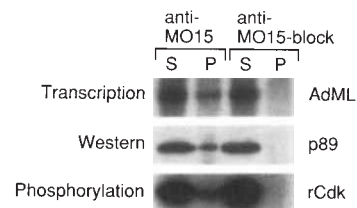


FIG. 3 TFIH is co-immunoprecipitated with MO15. TFIH transcription activity, the p89 (ERCC3) subunit of TFIH, and Cdk2 kinase activities are co-immunoprecipitated from the rat liver TFIH preparation with antibodies against MO15.

METHODS. Rat liver TFIH was incubated with anti-MO15 antibody and protein A agarose in 15 μ l of ELB²¹ with 1 mM DTT overnight at 4 °C as described in ref. 21. The mixture was spun in a microfuge for 10 s, and the supernatant (S) was removed. The pellet (P) was washed extensively with ELB with 1 mM DTT and resuspended in 15 μ l of buffer A¹⁶ containing 0.1 M KCl and 0.5 mg ml⁻¹ BSA. The anti-MO15-block reaction was done in parallel with an excess amount of the synthetic peptide used to generate the anti-MO15 antibody. The transcription, western blotting and kinase assays were done with equal amounts of supernatant and pellet suspensions as described in Fig. 1.

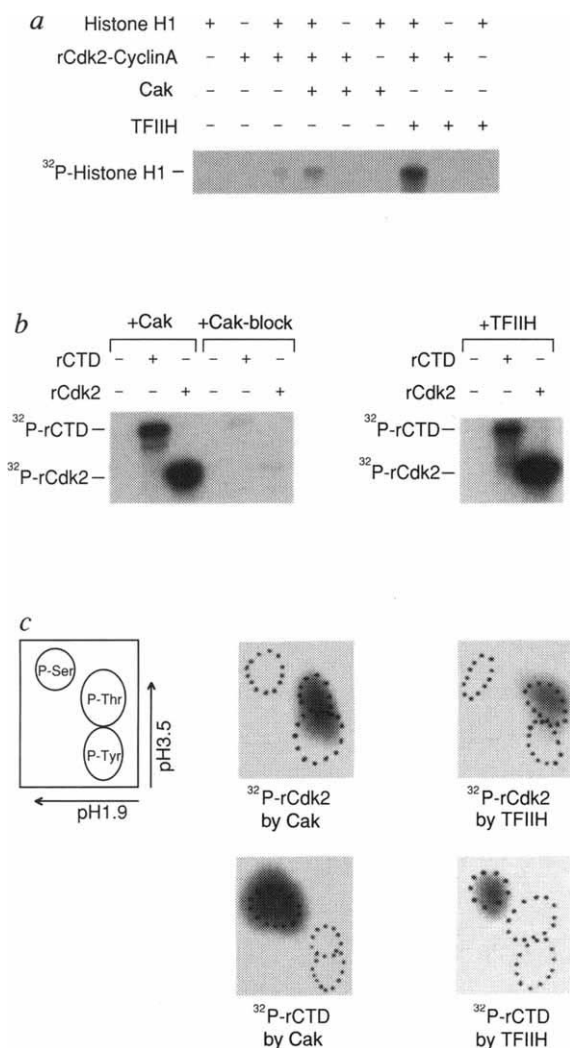


FIG. 4 Phosphorylation activities of rat liver TFIH and immunoprecipitated Cak. **a**, Highly purified rat liver TFIH can activate a recombinant Cdk2-cyclin A complex to phosphorylate histone H1. **b**, Immunoprecipitated Cak can phosphorylate recombinant CTD and Cdk2 fusion proteins. **c**, Amino acids phosphorylated by TFIH and Cak in Cdk2 and in the CTD.

METHODS. The kinase reaction mixtures shown in **a** were 15 μ l and contained 20 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM DTT, 0.2 mg ml^{-1} BSA, 10 μ M ATP, 2.5 μ Ci [γ -³²P] ATP. The mixtures were incubated at 28 °C for 10 min. To selected reactions were added histone H1 (Sigma), Cak immunoprecipitated from human C33a cells by anti-MO15 antibody²¹, highly purified rat liver TFIH, or a recombinant Cdk2-cyclin A complex prepared as described²¹. Histone H1 phosphorylation was analysed on 12% SDS-PAGE. Phosphorylation reactions in **b** with Cak immunoprecipitated from human C33a cells were as described²¹ with 0.5 μ g of GST-CTD (rCTD), 5 μ g of GST-Cdk2 (rCdk2) and anti-MO15 antibody. In the Cak-block preparation, an excess of MO15 synthetic peptide used to generate the anti-MO15 antibody was added to the immunoprecipitation reaction. Phosphorylation reactions by TFIH from rat liver were as described in Fig. 1 with 0.5 μ g of GST-CTD and 5 μ g of GST-Cdk2. Phosphorylated proteins were analysed on a 10% SDS polyacrylamide gel. For the analysis of phosphorylated amino acids shown in **c**, the GST-Cdk2 and GST-CTD fusion proteins phosphorylated by Cak and rat liver TFIH shown in **b** were transferred to PVDF membranes, isolated, hydrolysed with 6M HCl, and analysed by two-dimensional thin-layer electrophoresis as described³⁰.

ponents of TFIH are MO15 and cyclin H, providing a candidate for the enzyme responsible for phosphorylating the CTD and in this fashion controlling RNA polymerase II activity. Although Cak is clearly capable of phosphorylating the CTD *in vitro*, it is possible that it may not be the kinase that phosphorylates the CTD *in vivo*, but is rather an activator or regulator of one or more genuine CTD kinases.

The Cak enzyme has been reported to be responsible for the phosphorylation and attendant functional activation of at least three cyclin-dependent kinases, Cdc2 (ref. 19), Cdk2 (ref. 20) and Cdk4 (ref. 24). A direct proof of the role of TFIH-associated Cak in regulating RNA polymerase II function is difficult because reconstituted *in vitro* transcription systems appear not to depend on the presence of the CTD^{25,26} or its state of phosphorylation^{26,27}. With these caveats in mind, we suggest that the TFIH-associated Cak is indeed involved in regulating the activity of RNA polymerase II toward specific cellular genes. If this model is vindicated, it will mean that this single enzyme stands as a regulator of two of the most central apparatuses of the cell, the machinery responsible for gene expression and the cell-cycle clock.

Note added in proof: While this work was under review, MO15 and Kin28 were reported to be associated with mammalian and yeast TFIH preparations, respectively^{31,32}.

- Scafe, C. et al. *Nature* **347**, 491-494 (1990).
- Peterson, C. L., Kruger, W. & Herskowitz, I. *Cell* **64**, 1135-1143 (1991).
- Koleske, A. J., Buratowski, S., Nonet, M. & Young, R. A. *Cell* **69**, 883-894 (1992).
- Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. *Cell* **73**, 1361-1375 (1993).
- Koleske, A. J. & Young, R. A. *Nature* **368**, 466-469 (1994).
- Kim, Y.-J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. *Cell* **77**, 599-609 (1994).
- Cadena, D. L. & Dahmus, M. E. *J. Biol. Chem.* **262**, 12468-12474 (1987).
- Kolodziej, P. A., Woychik, N., Liao, S.-M. & Young, R. A. *Molec. cell. Biol.* **10**, 1915-1929 (1990).
- Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. *Cell* **67**, 1223-1330 (1991).
- Serizawa, H., Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7476-7480 (1992).
- Lu, H., Zawal, L., Fischer, L., Egly, J. M. & Reinberg, D. *Nature* **358**, 641-645 (1992).
- Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7356-7360 (1989).
- Serizawa, H., Conaway, R. C. & Conaway, J. W. *J. Biol. Chem.* **268**, 17300-17308 (1993).
- Schaeffer, L. et al. *Science* **260**, 58-63 (1993).
- Solomon, M. J., Lee, T. & Kirschner, M. W. *Molec. Biol. Cell* **3**, 13-27 (1992).
- Fisher, R. P. & Morgan, D. O. *Cell* **78**, 713-724 (1994).
- Mäkelä, T. P. et al. *Nature* **371**, 254-256 (1994).
- Gu, Y., Rosenblatt, J. & Morgan, D. O. *EMBO J.* **11**, 3995-4005 (1992).
- O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J. T. *Nature* **370**, 75-77 (1994).
- Matsuoka, M., Kato, J.-Y., Fisher, R. P., Morgan, D. & Sherr, C. J. *Molec. cell Biol.* **14**, 7265-7275 (1994).
- Buratowski, S. & Sharp, P. *Molec. cell. Biol.* **10**, 5562-5564 (1990).
- Li, Y. & Kornberg, R. D. *Proc. natn. Acad. Sci. U.S.A.* **91**, 2362-2366 (1994).
- Serizawa, H., Conaway, J. W. & Conaway, R. C. *Nature* **363**, 371-374 (1993).
- Conaway, R. C. & Conaway, J. W. *J. Biol. Chem.* **263**, 2962-2968 (1988).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- van der Geer, P., Luo, K., Sefton, B. M. & Hunter, T. in *Protein Phosphorylation* (ed. Hardie, D. G.) 31-59 (IRL, Oxford, 1993).
- Roy, R. et al. *Cell* **79**, 1093-1101 (1994).
- Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. *Cell* **79**, 1103-1109 (1994).

ACKNOWLEDGEMENTS. We initiated this study after learning that *S. cerevisiae* TFIH is associated with a CDK-related kinase (R. Kornberg, presented at the Whitehead Institute Symposium on Transcriptional Control, Cambridge, 3 October 1994). We thank D. Chao for doing the mass spectroscopy and are grateful to J. Parvin for the method of isolating functional TFIH through immunoprecipitation. This work was supported by the American Cancer Society (R.A.W.), by grants from the NIH (to R.A.Y., R.C.C. and J.W.C.) and by funds provided to the Oklahoma Medical Research Foundation by the H.A. and Mary K. Chapman Charitable Trust (R.C.C. and J.W.C.). T.P.M. is a fellow of the Human Frontier Science Program Organization. R.A.W. is an American Cancer Society Research Professor.

Received 9 November 1994; accepted 2 February 1995.

- Corden, J. L. *Trends biochem. Sci.* **15**, 383-387 (1990).
- Young, R. A. *Rev. Biochem.* **60**, 689-715 (1991).
- Dahmus, M. E. & Dynan, W. S. in *Transcriptional Regulation* 109-128 (Cold Spring Harbor Laboratory Press, New York, 1992).
- Allison, L. A. & Ingles, C. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2794-2798 (1989).

Cdk-activating kinase complex is a component of human transcription factor TFIH

Ramin Shiekhattar, Fred Mermelstein,
Robert P. Fisher*, Ronny Drapkin,
Brian Dynlacht†, Holly C. Wessling*,
David O. Morgan* & Danny Reinberg‡

Howard Hughes Medical Institute, Department of Biochemistry,
Robert Wood Johnson Medical School, University of Medicine and
Dentistry of New Jersey, 675 Hoes Lane, Piscataway,
New Jersey 08854-5635, USA

* Department of Physiology, University of California, San Francisco,
California 94143-0444, USA

† Laboratory of Molecular Oncology, The MGH Cancer Center,
Building 149, 13th Street, Charlestown, Massachusetts 02129, USA

‡ To whom correspondence should be addressed

TRANSCRIPTION factor IIF (TFIIF) contains a kinase capable of phosphorylating the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNAPII)¹⁻³. Here we report the identification of the Cdk-activating kinase (Cak) complex (Cdk7 and cyclin H) as a component of TFIIF after extensive purification of TFIIF by chromatography. We find that affinity-purified antibodies directed against cyclin H inhibit TFIIF-dependent transcription and that both cyclin H and Cdk7 antibodies inhibit phosphorylation of the CTD of the largest subunit of the RNAPII in the preinitiation complex. Cak is present in at least two distinct complexes, TFIIF and a smaller complex that is unable to phosphorylate RNAPII in the preinitiation complex. Both Cak complexes, as well as recombinant Cak, phosphorylate a CTD peptide. Finally, TFIIF was shown to phosphorylate both Cdc2 and Cdk2, suggesting that there could be a link between transcription and the cell cycle machinery.

Immunoaffinity purification using antibodies directed against the 62K subunit of TFIIF was used to identify any CTD kinase

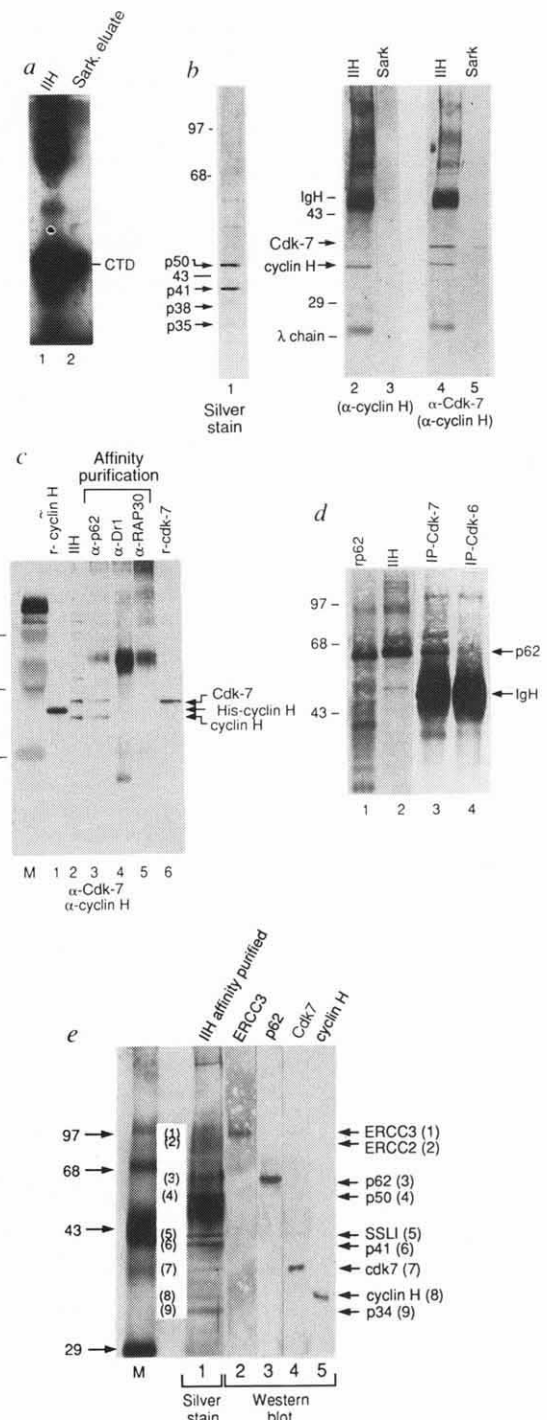


FIG. 1 *a*, The Sarkosyl eluate contains CTD kinase activity. lane 1, TFIIF (phenyl-Superose, 230 ng); lane 2, Sarkosyl eluate (5 μ l). Labelled products near the origin of the gel in lane 1 represent phosphorylation of other proteins in the phenyl-Superose preparation. The phosphorylation assay has been described¹. *b*, Silver-stained and western blot analysis of polypeptides dissociated from immunoaffinity-purified TFIIF by the Sarkosyl wash. lane 1, silver stain of the peak Sarkosyl fraction. Arrows indicate enriched polypeptides in this fraction. Lanes 2–5, western blot of Cdk7 and cyclin H in an immunopurified TFIIF (lanes 2 and 4) and Sarkosyl eluate of the p62 immunoaffinity purification (lanes 3 and 5). Lanes 2 and 3 of the blot were incubated with cyclin H antibodies only. Lanes 4 and 5, western blot incubated with both cyclin H and Cdk7 antibodies. Arrows to the left indicate cyclin H and Cdk7. IgH represents the immunoglobulin heavy chain; λ chain is the immunoglobulin light chain. Molecular weight markers (in thousands) are indicated at the left. *c*, Identification of Cdk7 and cyclin H as components of TFIIF using immunoaffinity chromatography with p62 antibodies coupled to protein A-agarose. Lanes: 1, recombinant hexahistidine-tagged cyclin H (50 ng); 2, TFIIF (phenyl-Superose 2.3 μ g); 3, anti-p62 eluate; 4, anti-Dr1 eluate; 5, anti-RAP30 eluate; 6, recombinant Cdk7 (50 ng). Arrows to the right indicate the position of Cdk7, cyclin H, and His-tagged cyclin H. Molecular weight markers are indicated at the left of the gel. Purification was as already described, except that the column was eluted with 100 mM glycine, pH 2.6, after the low-salt wash. *d*, Cdk7 antibodies, but not Cdk6 antibodies, immunoprecipitate the 62K subunit of TFIIF. Lanes: 1, recombinant p62 (100 ng); 2, TFIIF (phenyl-Superose, 2.6 μ g); 3 and 4, as indicated. *e*, Silver-stained and western blot analysis of TFIIF was resolved on the same gel and directly compared with known TFIIF subunits. Lanes: 1, silver staining of immunopurified TFIIF; 2–5, western

blots of TFIIF (phenyl-Superose, 2.3 μ g) using the antibodies indicated (top). Arrows on the right depict known TFIIF polypeptides. Numbers to the right of the figure correspond to the polypeptides indicated by arrows on the left of the gel. Molecular weight markers are shown by arrows on the left. The same polypeptides are also shown in the silver stain of Fig. 4.

METHODS. In *b*, polyclonal rabbit antibodies (2 mg) directed against the p62 subunit of TFIIF were covalently crosslinked to 1 ml protein A-agarose using 20 mM dimethylpimelidate as described¹³. The resin was incubated for 90 min at 4 °C with nuclear extract (4 ml, 9 mg ml⁻¹) prepared as described¹⁴ and dialysed against buffer containing 20 mM Tris-HCl, pH 7.9, 0.2 mM EDTA, 50 mM NaF, 100 μ M NaVO₃, 100 mM KCl, 1% N-P40 and 1 mM PMSF. Immunoabsorbed TFIIF was then washed with 10 ml of this buffer containing 1 M KCl and 1% N-P40. This was followed by a low-salt wash (100 mM KCl; 10 ml) and elution in low-salt buffer containing 0.05% Sarkosyl; remaining bound material was washed again in low-salt buffer and eluted with 0.1 M glycine-HCl, pH 2.6.

activity associated with this factor. In an attempt to dissociate the polypeptide(s) containing kinase activity, the immuno-adsorbed TFIH complex was treated with the ionic detergent Sarkosyl. The remaining immuno-adsorbed TFIH polypeptides were eluted at low pH (a glycine wash). Analysis of the Sarkosyl eluate revealed the presence of a kinase activity capable of phos-

phorylating a tetra-heptapeptide repeat (YSPTSPS) of the CTD complex (Fig. 1a), but not the CTD of RNAPII in the preinitiation complex (data not shown). Silver-stain analysis of this fraction indicated the presence of polypeptides ranging in size from 35K to 50K (Fig. 1b, lane 1). Cdk-related kinases phosphorylate the CTD^{4,6}, and we observed that the molecular mass of the

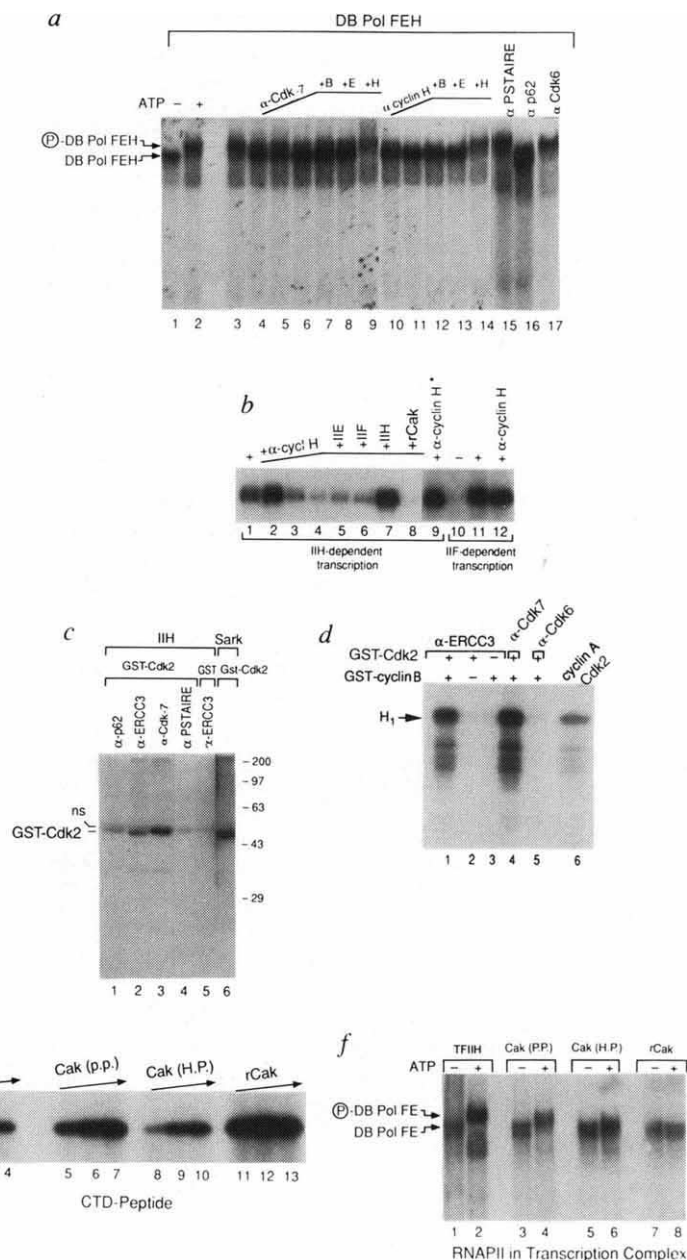
FIG. 2 a, Affinity-purified antibodies against Cdk7 and cyclin H inhibits the phosphorylation of RNAPII in the pre-initiation complex. All antibodies were incubated with TFIH for 1 h at 4 °C before addition to the binding reaction (lanes 4–17). Lanes: 1 and 2, complete complex in the absence and the presence of 1 mM ATP, respectively; 3–17, complete complex in the presence of 1 mM ATP (in lane 3, TFIH was incubated in the absence of any antibody, and in lanes 4–6, TFIH was incubated with increasing concentrations of affinity-purified Cdk7 antibodies (1, 2, 3 µg, respectively); 10 and 11, TFIH incubated with 150 and 250 ng affinity purified cyclin H antibodies, respectively; 7–9, incubation of TFIH with 3 µg Cdk7 antibodies, after which 2 × TFIIB (20 ng, lane 7), 2 × TFIIE (lane 8), or 2 × TFIIF (270 ng, lane 9) was added to the reaction; 12–14, products of reactions in which TFIH was incubated with 250 ng cyclin H antibodies, after which 2 × TFIIB (20 ng, lane 12), 2 × TFIIE (lane 13), or 2 × TFIIF (270 ng, lane 14) was added; 14–17, incubation of TFIH with 600 ng PSTAIRE, p62 or Cdk6 antibodies. DNA-binding reactions have been described¹. The faster-migrating complex contains unphosphorylated RNAPII and is designated (DB)PolFEH; the slower-migrating complex contains phosphorylated RNAPII (P-DB)PolFEH).

b, Cyclin H antibodies inhibit TFIH-dependent transcription but not TFIIF-dependent transcription. Lane 1 shows the product of a complete TFIH-dependent reaction; lanes 2–4 contain partially purified TFIH incubated with increasing amounts of affinity-purified cyclin H antibodies (50, 150 and 250 ng, respectively). In the remaining lanes, TFIH was first incubated with 250 ng cyclin H antibodies, after which the complete reaction mixture, together with saturating amounts of recombinant TFIIE (lane 5), TFIIF (lane 6), TFIH (phenyl-Superose, lane 7) or rCAK (lane 8), was added to each tube. Lane 9, transcription reaction in which 250 ng cyclin H antibodies were added to a complete transcription mix; lanes 10 and 11, transcription reactions in the presence and the absence of TFIIF, respectively; lane 12, incubation of TFIIF with cyclin H antibodies for 1 h before addition to the transcription reaction. Transcription reactions have been described¹⁵.

c, Phosphorylation of Cdk2 by TFIH. TFIH or CAK complexes were immunoprecipitated using p62, ERCC3, Cdk7 or PSTAIRE antibodies as described¹⁰, except the beads were used in a kinase assay. Lanes 1–4, phosphorylation of GST-Cdk2 by immunoprecipitated complexes indicated (top); lane 5, control, using GST as the substrate; lane 6, Sarkosyl wash of the p62 immuno-affinity-purified TFIH phosphorylates GST-Cdk2. ns, Nonspecific band. Phosphorylation was assayed as described⁸. ERCC3 monoclonal antibodies were produced in collaboration with Austral Biological-Bios, Chile.

d, Histone H1 phosphorylation by Cdk2 in the absence or presence of cyclin B. Lanes 1–5, complexes immunoprecipitated using the antibodies shown (top); lane 6 is histone H1 phosphorylation by cyclin A-Cdk2. Activation was assayed as described⁸.

e, Multiple Cak complexes can phosphorylate the CTD peptide. CTD peptide was incubated with different Cak-containing complexes as indicated (top). The amount of Cak complexes used in the assays in e and f were: TFIH (2 µl of fraction 27 of gel filtration step (Fig. 4); highly purified Cak (H.P.) corresponding to a fraction from a phosphocellulose column (40 ng)⁸; partially purified Cak (P.P.) is a fraction from the S200 sizing column (1 µg)⁸. Recombinant Cak is composed of recombinant His-cyclin H and baculovirus-expressed Cdk7 (50 ng Cdk7 and 100 ng cyclin H). f, Gel mobility-shift assay showing the effect of different Cak complexes on the phosphorylation of RNAPII in the preinitiation complex.



proteins in our Sarkosyl eluate resembled that of Cdk-related kinases (M_r 30K to 40K). We therefore analysed purified TFIIF fractions for the presence of cell-cycle-related kinases and found that TFIIF is devoid of any immunoreactive material representing the kinases Cdk1-Cdk6, and PCTAIRE 1-3 (kinases with homology to Cdks) (data not shown). However, human Cak, composed of Cdk7/MO15 and cyclin H, contains polypeptides ranging in size from 35K to 40K^{7,8}. Western blot analysis of highly purified TFIIF, affinity-purified TFIIF, or Sarkosyl eluate revealed the presence of material immunoreactive with Cdk7 and cyclin H in all three preparations (Fig. 1b, lanes 2-5 and Fig. 1c, lanes 2 and 3). The presence of Cak polypeptides in these fractions was specific, as eluates derived from Dr₁ or RAP30 affinity columns were devoid of Cdk7 and cyclin H (Fig. 1b, lanes 4 and 5). Furthermore, immunoprecipitation of TFIIF using Cdk7 (Fig. 1d, lane 3) or cyclin H (R.F. and D.M., unpublished observations) antibodies revealed P62 immunoreactive material. This p62 immunoreactivity was not evident with Cdk6 antibodies (Fig. 1d; compare lanes 3 and 4). Silver-stained and western blots of the p62-immunopurified TFIIF complex is shown in Fig. 1e. In addition to Cdk7 and cyclin H, other cloned TFIIF polypeptides and putative subunits are present in this preparation (arrows in Fig. 1e). Thus we have immunological evidence for the presence of Cdk7 and cyclin H in the TFIIF complex.

We previously demonstrated that when the complete transcription initiation complex (DBPolFEH) is incubated with ATP, the mobility of the complex is retarded as a result of CTD phosphorylation by TFIIF in a gel mobility shift assay (refs 1,

6, 9; Fig. 2a, lanes 1 and 2). Incubation of TFIIF with increasing amounts of affinity-purified antibodies against Cdk7 (lanes 4-6), cyclin H (lanes 10 and 11) or p62 (lane 16) before its addition to the reaction resulted in a change in migration of the complex: the complex now co-migrated with the complete complex formed in the absence of ATP (lane 1). Inhibition of phosphorylation was overcome by addition of excess TFIIF (lanes 9 and 14) but not by excess TFIIB or TFIIE (lanes 7, 12 and 8, 13, respectively). Furthermore, antibodies directed against the PSTAIRE motif or Cdk6 failed to inhibit the ATP- and TFIIF-dependent shift (lanes 15 and 17).

Incubation of TFIIF with increasing concentrations of cyclin H antibodies before its addition in the reconstituted transcription assay resulted in inhibition of transcription (Fig. 2b, lanes 3 and 4). At low concentrations, cyclin H antibodies had a mild stimulatory effect on transcription (Fig. 2b, lane 2). The dose-dependent inhibition was specific as it was overcome by addition of excess TFIIF (lane 7), but not by excess TFIIE or TFIIF (lanes 5 and 6). This inhibition was not due to a nonspecific inactivation of TFIIF by the antibodies, as incubation of these antibodies with TFIIF did not affect transcription (Fig. 2b, lanes 10-12). Interestingly, incubation of TFIIF with antibodies against Cdk7 was without effect in transcription (data not shown). Inhibition of transcription by antibodies against cyclin H may result from exclusion of TFIIF from the preinitiation complex, consistent with the lack of effect when cyclin H antibodies were added after the preinitiation complex was formed (Fig. 2b, lane 9). As TFIIF has a dual role in transcription and nucleotide-excision repair^{10,12}, the effect of Cdk7 antibodies was

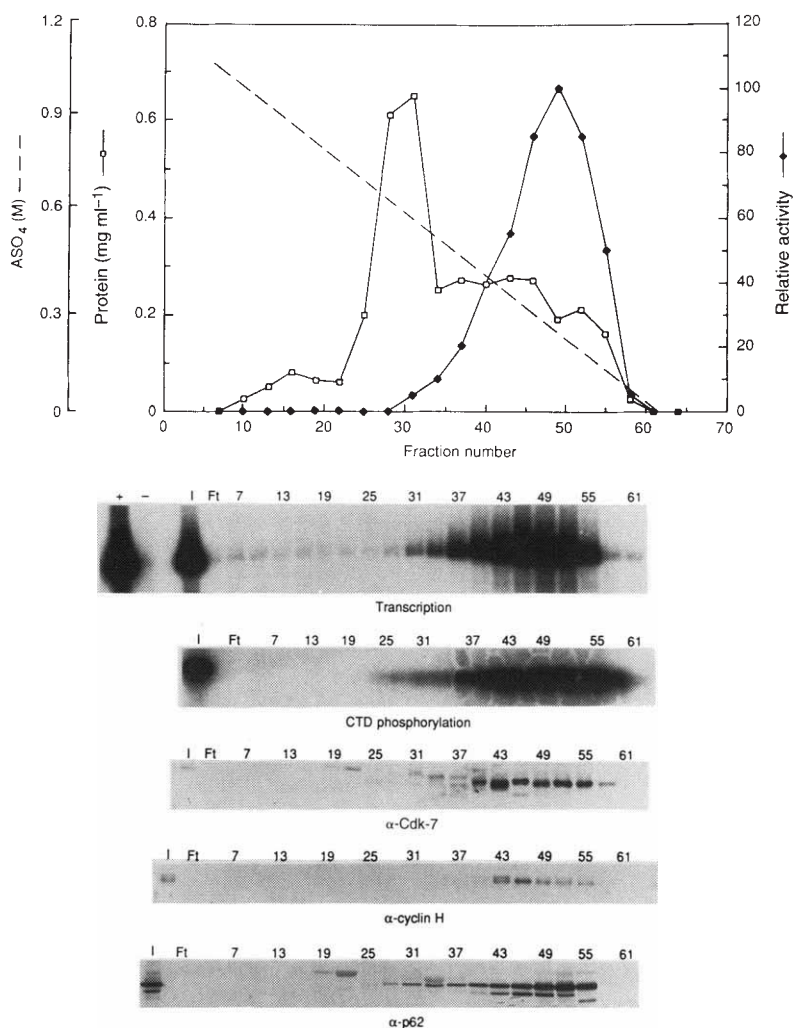


FIG. 3 TFIIF transcription activity copurifies with p62, cyclin H and Cdk7 polypeptides, and with CTD-peptide kinase activity on phenyl-Superoose chromatography. METHODS. TFIIF was purified from 5.5 g HeLa nuclear extracts and fractionated as described through the fourth chromatographic step¹⁰. Transcription reactions were performed as described¹⁰ except that TFIIF was omitted from the general transcription factor mix and substituted with the input (I) to the column, the flow-through (Ft), or the fractions indicated at the top of each lane.

tested in a reconstituted nucleotide-excision repair reaction where, in contrast to their effect in transcription, they were inhibitory (D. Mu, R.S., R.D., A. Sancar and D.R., unpublished observation). It is not clear why Cdk7 antibodies inhibit phosphorylation of the CTD as well as nucleotide-excision repair functions of TFIIH, while they have no effect in transcription. The epitope recognized by the antibodies could be critical for one function and not the other. Nonetheless, the inhibition of phosphorylation, transcription and nucleotide-excision repair by antibodies directed against Cak components demonstrates that Cak polypeptides are integral components of the TFIIH complex.

We next examined whether the TFIIH complex has Cak activity. Analysis of immunopurified complexes using p62, ERCC3 or Cdk7 antibodies indicated that all three complexes phosphorylate the glutathione-S-transferase fusion protein GST-Cdk2 (Fig. 2c). The extent of Cdk2 phosphorylation varied among the three immunopurified complexes, with Cdk7 being the most active. This may be due to the ability of Cdk7 antibodies to precipitate other Cak-containing complexes as well as TFIIH (see below). The lower activity observed with p62 antibodies may be due to a partial inhibition of the kinase activity by these antibodies (Fig. 2a, lane 16). A PSTAIRE immunoprecipitated complex was devoid of any activity (Fig. 2c, lane 4). No kinase activity was detected when GST alone was used as a substrate (Fig. 2c, lane 5). The Sarkosyl eluate of a complex immunopurified using p62 antibodies could also phosphorylate Cdk2 (Fig. 2c, lane 6). We also tested for histone H1 activation by immunopurified TFIIH complexes using ERCC3 or Cdk7 antibodies. Figure 2d shows that a TFIIH complex immunopurified using ERCC3 as well as Cdk7 antibodies can activate Cdk2 in a cyclin-dependent manner (compare lane 2 with lanes 1 and 4). Cdk6 immunopurified complexes were devoid of any Cak activity (Fig. 2d, lane 5); results were similar when Cdc2 was used as substrate (data not shown). Together, these results demonstrate that Cak, in the context of the TFIIH complex, can activate Cdk-related kinases.

We next analysed the CTD-phosphorylating activity of recombinant Cak. Unlike TFIIH, which could phosphorylate the CTD of RNAPII in the preinitiation complex, rCak could not (Fig. 2f; compare lanes 1 and 2 with 7 and 8); however, it could efficiently use the CTD peptide as a substrate (Fig. 2e, lanes 11–13). Furthermore, we found that although either a partially or highly purified preparation of native Cak could act on the CTD peptide (Fig. 2e, lanes 5–10), only the less pure preparation could phosphorylate the CTD in the preinitiation complex (Fig. 2f, lanes 3–6). These results not only show that Cdk7 and cyclin H

polypeptides (Cak) recognize the CTD heptapeptide sequence as a substrate, but they also suggest that there could be at least two different Cak-containing complexes, TFIIH and a complex that is unable to phosphorylate the CTD of RNAPII. In support of this, we found that neither highly purified native Cak nor rCak could replace TFIIH in the transcription assay (data not shown).

In an attempt to isolate different Cak-containing complexes, we assayed fractions from the different steps of the TFIIH purification for TFIIH-dependent transcription, CTD phosphorylation and Cdk2 activation (measured by phosphorylation of histone H1 activity), as well as for the presence of p62 (a core subunit of TFIIH), cyclin H and Cdk7 polypeptides by western blot. This analysis revealed that transcription, as well as CTD-peptide kinase activity copurified with p62, cyclin H and Cdk7 immunoreactivity at the fourth step of purification (phenyl-Superose chromatography; Fig. 3). But further fractionation of

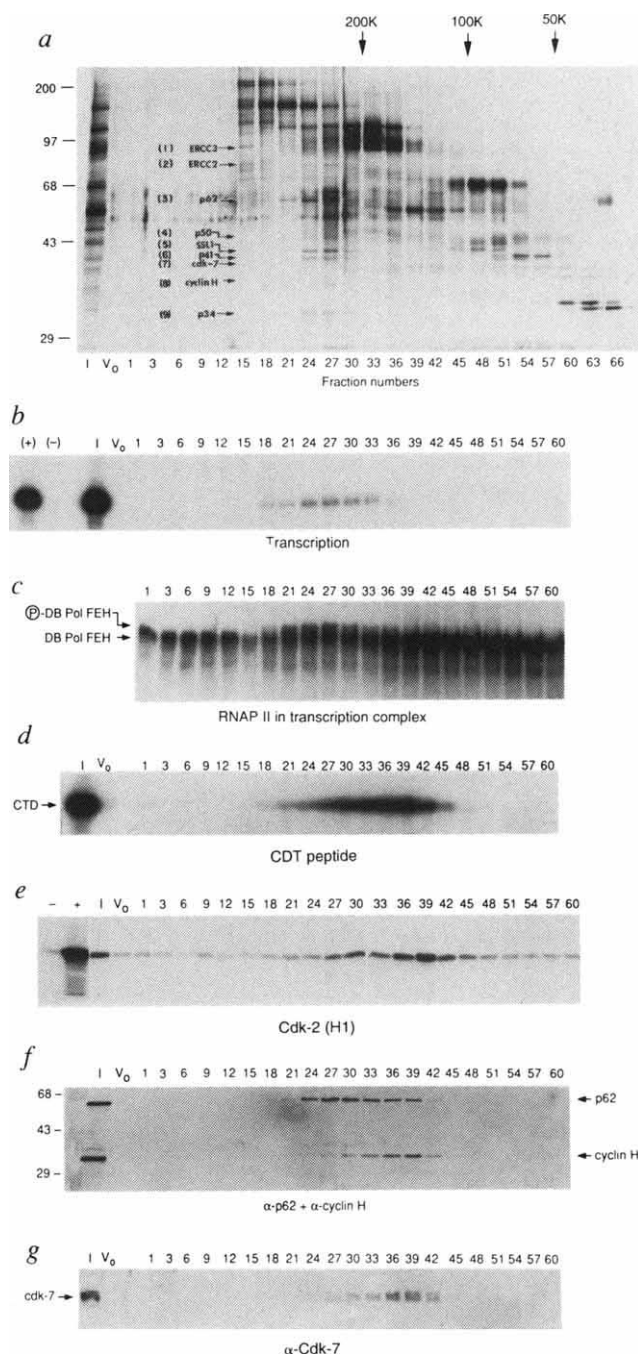


FIG. 4 Fractionation of TFIIH on gel filtration chromatography reveals the presence of two Cak-containing complexes. **a**, Silver staining of the fractions eluted from the gel-filtration column. **I**, input to the column; **V₀**, void volume. Arrows at the top of the gel represents the elution of molecular weight markers. Fraction numbers are indicated at the bottom. Migration of molecular weight protein standards is indicated on the left. The subunits of the TFIIH complex are indicated by arrows. The numbers indicated to the left of the TFIIH subunits correspond to the polypeptides observed in the affinity-purified TFIIH shown on Fig. 1e. **b**, TFIIH-dependent transcription analysis of gel-filtration column. **c**, Phosphorylation of RNAPII within the preinitiation complex. **d**, Phosphorylation of CTD tetra-heptapeptide. **e**, Cak activity as measured by activation of Cdk2 histone H1 activity. **f**, Western blot analysis for TFIIH (p62) and cyclin H polypeptides. **g**, Western analysis for Cdk7 protein. **METHODS**. **a**, The active TFIIH fractions from the phenyl-Superose column (Fig. 3) were pooled and a portion (1.2 mg) fractionated on a Pharmacia Mono S 5/5 column. The transcriptionally active fractions were pooled (800 µg) and fractionated on a Pharmacia Superdex-200 16/60 column in the presence of 1.0 M KCl and 0.01% N-P40. Fractions were subjected to SDS-PAGE and silver staining. Activation of Cdk2 measured as phosphorylation of histone H1 was done as described⁸. 1 µg Cdk2 bound to protein A-Sepharose was incubated with 10 µl of each fraction.

TFIIH on a gel filtration column apparently separated two distinct Cak-containing complexes. The larger complex (best represented by fractions 24 and 27) contained both TFIIH-dependent transcription (Fig. 4b) and kinase activities (measured as phosphorylation of the CTD of RNAPII within the preinitiation complex; Fig. 4c). Both activities were weak, perhaps as a result of dilution by the sizing-column fractionation (fractions 18–39, peaking at fractions 24 and 27; Fig. 4b, c). These activities coeluted with polypeptides containing TFIIH (Fig. 4a; also western blot of p62 in Fig. 4f). However, the peak of CTD-peptide phosphorylation as well as Cak activity (Cdk2 phosphorylation of histone H1) was shifted to fraction 39, away from the peak of transcription. Similarly, western blot analysis of Cdk7 and cyclin H polypeptides indicated that the peak of Cdk7 and cyclin H immunoreactivity coincided with that of Cdk2 activation and CTD-peptide phosphorylation (Fig. 4f, g). Cdk7 and cyclin H immunoreactivity extended from fraction 18 to 45 (Fig. 4f, g), indicating that all fractions containing transcription activity also contained Cdk7- and cyclin-H-immunoreactive material. These results suggest the presence of at least two distinct Cak-

containing complexes, a larger transcriptionally active TFIIH ($M_r > 200K$) complex and a smaller complex unable to phosphorylate RNAPII in the preinitiation complex. The smaller Cak complex(es) may lack components of TFIIH that tether the factor to the preinitiation complex. Alternatively, Cak may acquire a different substrate specificity upon interacting with the other TFIIH subunits. Also, Cak may be activating another CTD kinase within the transcription complex that is capable of phosphorylating the polymerase.

TFIIH phosphorylation of Cdk-related kinases raises the question of whether TFIIH is involved in cell-cycle regulation. Our results indicate that most of the Cak is in a complex(es) distinct from TFIIH, and that the direct involvement of TFIIH in cell-cycle progression is unlikely. However, TFIIH, by phosphorylating Cdc2 or Cdk2, may play a role in the modulation of transcription and DNA repair during the cell cycle.

Note added in proof: While this Letter was under review, a report appeared demonstrating that MO15 is a component of human TFIIH (ref. 16), and another indicated that KIN28 is the MO15-homologue in yeast TFIIH (ref. 17). □

Received 7 December 1994; accepted 6 February 1995.

1. Lu, H., Zawal, L., Fisher, L., Egly, J. M. & Reinberg, D. *Nature* **358**, 641–645 (1992).
2. Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. D. *Cell* **67**, 1223–1230 (1991).
3. Serizawa, H., Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7476–7480 (1992).
4. Cisek, L. J. & Corden, J. L. *Nature* **339**, 679–684 (1989).
5. Cisek, L. J. & Corden, J. L. *Meth. Enzym.* **200**, 301–325 (1991).
6. Zawal, L., Lu, H., Cisek, L. J., Corden, J. L. & Reinberg, D. *Cold Spring Harbor Symp. quant. Biol.* **58**, 187–198 (1993).
7. Makela, T. P. et al. *Nature* **371**, 254–257 (1994).
8. Fisher, R. P. & Morgan, D. O. *Cell* **78**, 713–724 (1994).
9. Lu, H., Flores, O., Weinmann, R. & Reinberg, D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10004–10008 (1991).
10. Drapkin, R. et al. *Nature* **368**, 769–772 (1994).
11. Van Vuuren, A. J. et al. *EMBO J.* **13**, 1645–1653 (1994).

12. Wang, Z. et al. *Nature* **368**, 74–75 (1994).
13. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988).
14. Dignam, D. J., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1485 (1983).
15. Flores, O., Lu, H. & Reinberg, D. *J. biol. Chem.* **267**, 2786–2793 (1992).
16. Roy, R. et al. *Cell* **79**, 1093–1101 (1994).
17. Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. *Cell* **79**, 1103–1109 (1994).

ACKNOWLEDGEMENTS. R.S. and F.M. contributed equally to this work. We thank M. Solomon, E. Liu, C. Sherr, H. Zhang, D. Beach, T. Makela and R. Weinberg for reagents; O. Aprelikova for Cdk7 antibodies; E. Harlow for fostering our collaboration with B.D.; members of our laboratories, R. Young and T. Lagrange, for discussion; M. Sheldon for help in preparing the manuscript and C. Selby and T. Lagrange for critical comments on it. R.S. is supported by an NIH postdoctoral grant; B.D. by the Damon Runyon cancer fund; and R.D. by an NIH predoctoral grant. This work was supported by grants from the NIH and the Howard Hughes Medical Institute to D.R. D.R. was a recipient of an American Cancer Society Faculty Research Award.

Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops

Kenneth J. McDowall*, Vladimir R. Kaberdin†, Se-Wei Wu†, Stanley N. Cohen*† & Sue Lin-Chao*†

* Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

† Institute of Molecular Biology, Academia Sinica, Nankang 11529, Taipei, Taiwan, Republic of China

THE enzyme RNase E (ref. 1) cuts RNA at specific sites within single-stranded segments^{2–6}. The role of adjacent regions of secondary structure in such cleavages is controversial^{7–10}. Here we report that 10–13-nucleotide oligomers lacking any stem-loop but containing the RNase E-cleaved sequence of RNA I^{11–13}, the anti-sense repressor of replication of ColE1-type plasmids, are cut at the same phosphodiester bond as, and 20 times more efficiently than, RNA I. These findings indicate that, contrary to previous proposals^{8,9}, stem-loops do not serve as entry sites for RNase E, but instead limit cleavage at potentially susceptible sites. Cleavage was reduced further by mutations in a non-adjacent stem-loop, suggesting that distant conformational changes can also affect enzyme access. Modulation of RNase E cleavages by stem-loop regions and to a lesser extent by higher-order structure may explain why this enzyme, which does not have stringent sequence specificity^{6,12,13}, cleaves complex RNAs at a limited number of sites.

RNase E cleaves RNAs of known secondary structure (such as RNA I, 9S ribosomal RNA, S20 messenger RNA) within single-stranded regions rich in A + U nucleotides^{6,8,11} (see Fig. 1a for pBR322 RNA I). Mutational modification of the single-stranded segment bracketing the RNase E cleavage site at the 5' end of RNA I encoded by the pBR322 or pACYC184 plasmid indicates that there is no simple relationship between the order of nucleotides and the phosphodiester bond cleaved^{11,12}. Although nearby regions of secondary structure can affect RNase E cleavages^{6–10,12}, the actual role of such regions has been unclear. Base pairing of nucleotides at the 5' terminus of RNA can interfere with RNase E cleavage *in vivo*⁷. However, internally located stem-loops positioned 5' to RNase E sites have been proposed to promote cleavage, possibly by enabling loading of the enzyme onto the substrate⁸. Others have suggested that stem-loops 3', rather than 5', to the site of cleavage are necessary for the entry of RNase E, which then scans the upstream segment for a sequence that can be cleaved⁹. Still another proposal is that stem-loops adjacent to RNase E cleavage sites may function as anchors to maintain the cleaved region in a single-stranded conformation^{8,10}.

To learn whether a stem-loop is in fact required as an entry site for RNase E, we determined whether short oligoribonucleotides that lack the potential to form stem-loops can serve as substrates. As shown in Fig. 2, RNase E purified from *Escherichia coli* by immobilized metal affinity chromatography, cleaved two decaribonucleotides, BR10 (5'-ACAGAUUUUG) and ACYC10 (5'-ACAAGUUUUG), which correspond to the 5' ends of RNA I from pBR322 and pACYC184, respectively, at the same phosphodiester bonds found previously to be cleaved by RNase E in the corresponding native RNA I species^{11,13}. Furthermore, affinity-purified RNase E also cleaved a tridecaribonucleotide, BR13 (5'-GGGACAGAUUUUG; Fig. 2) identical in sequence to the 5' end of GGG.RNA I¹⁴, a T7 RNA

† To whom correspondence should be addressed.

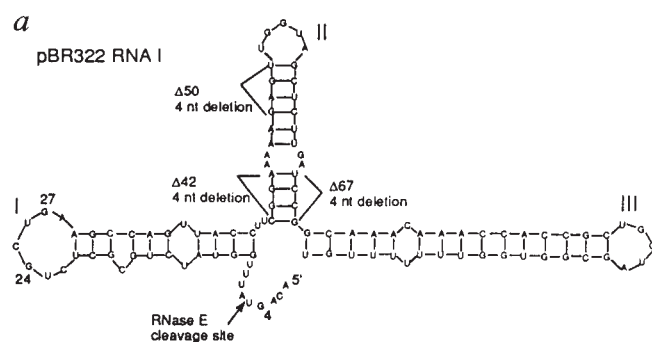
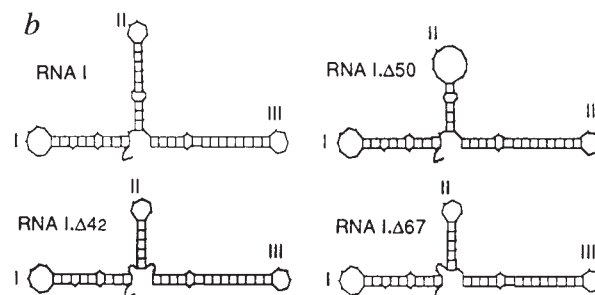


FIG. 1 Structure of pBR322 RNA I and RNA I variants used in this study. **a**, Secondary structure of pBR322 RNA I as determined previously^{14,18}. The positions of tetranucleotide deletions in stem-loop II are indicated and an arrow marks the site of RNase E cleavage^{11,12}. **b**, Secondary



structures of RNA I.Δ42, RNA I.Δ50 and RNA I.Δ67, as predicted by the FOLD computer program¹⁹ and confirmed by RNase T₁ digestion analysis (Fig. 4a).

polymerase-synthesized pBR322 RNA I variant, at the same site cleaved in the corresponding full-length RNA (Fig. 3). These results show that the primary sequences of these single-stranded oligoribonucleotides are sufficient to direct site-specific cleavage by RNase E.

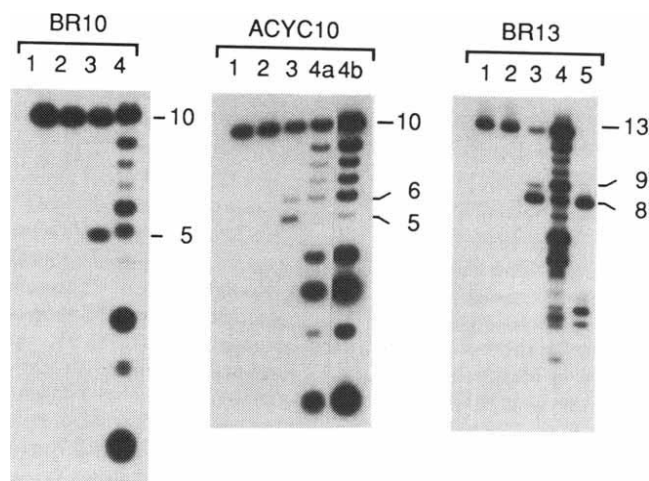
We also observed that full-length RNA I substrates were cleaved much less efficiently than the corresponding oligoribonucleotides. Under precisely the same experimental conditions, GGG.RNA I was cut by RNase E at only 5% the rate observed for oligoribonucleotide BR13 (Fig. 3). The less efficient cleavage of full-length RNA I indicates that regions of secondary structure close to, and downstream from, the cleavage site negatively affect the cleavage rate. The oligoribonucleotide BR13 was also

cleaved 20 times faster than GGG.RNA I when both substrates were incubated together, and when the enzyme concentration was increased by fivefold and transfer RNA was omitted from the reactions. Additionally, we found that a previously described RNase E preparation¹² and a renatured RNase E-derived peptide consisting of only the N-terminal domain of the protein (amino acids 1 to 498) and purified to >95% homogeneity by affinity chromatography and denaturing preparative gel electrophoresis also cleaved oligoribonucleotide BR13 about 20 times more efficiently than GGG.RNA I (data not shown).

Other experiments showed that mutations in RNA I distant from the RNase E site can further reduce cleavage by another 50% to 80% while not affecting the stem-loop adjacent to the

FIG. 2 Site-specific cleavage of oligoribonucleotides by RNase E. Each 5'-labelled oligoribonucleotide (1–5 pmol) was incubated separately with 15 ng of affinity-purified RNase E and the reaction products were separated in a 20% polyacrylamide sequencing gel. Lane 1, 5'-labelled oligoribonucleotide (gel purified); lanes 2 and 3 as lane 1, except that the oligoribonucleotide was incubated in cleavage buffer without or with RNase E, respectively, for 30 min; lane 4, 1-nt ladder generated by partial digestion of the oligoribonucleotide with mung bean nuclease. Numbers at the right of the autoradiographs indicate the nucleotide length of the principal RNA species. 4a and 4b of the middle panel are short and long exposures, respectively, of the same sample. The sequences of, and major sites of RNase E cleavage within (vertical arrow), oligoribonucleotides BR10, ACYC10 and BR13 are 5'-ACAGU[^]AUUUG, 5'-ACAAG[^]UUUUG and 5'-GGGACAGU[^]AUUUG, respectively. Lanes 5 of the BR13 panel contains 5'-labelled octamer GGGACAGU (not gel-purified).

METHODS. Full-length RNase E tagged at the N terminus by histidine (His) residues was overexpressed *in vivo* from pFUS1500 (supplementary information), a pET16b-based construct encoding the full-length *rne* gene, and purified by non-denaturing immobilized metal affinity chromatography²⁰ (IMAC) as described in the pET System Manual (Novagen), except that all of the chromatography buffers contained 0.1% Triton X-100 (Sigma). His-tagged RNase E was eluted from the IMAC column using 200–300 mM imidazole; concentrated to a volume less than 250 μ l using a Centricon-10 (Amicon); dialysed against 1 litre of 50 mM Tris-Cl (pH 7.5), 20% glycerol, 1 mM EDTA, 500 mM NaCl, and 0.5% Triton X-100; dithiothreitol was added to final concentration of 1 mM; and stored at -70°C until used. About 50% of the preparation, which cleaved 9S RNA and pBR322 RNA I highly specifically at previously mapped RNase E sites, contained full-length RNase E protein as determined by Coomassie blue staining of samples run in SDS-polyacrylamide gels. The oligoribonucleotides were synthesized using standard phosphoramidate chemistry (Integrated DNA Technologies), labelled at the 5' end using T4 polynucleotide kinase, and gel purified as described previously for oligodeoxynucleotide primers^{1,3}. Cleavage reactions: substrates were incubated at 30°C with affinity-purified RNase E in 50 μ l of 20 mM Tris-HCl (pH 8.0), 5 mM MgCl_2 , 100 mM NaCl, 5% glycerol, 0.1% Triton X-100, 0.1 mM dithiothreitol, containing yeast tRNA at a concentration of 80 $\mu\text{g ml}^{-1}$.



segment cleaved. We constructed three RNA I variants ($\Delta 42$, $\Delta 50$, and $\Delta 67$; Fig. 1) containing stem-loop II tetranucleotide deletions (Fig. 1a) that were predicted not to change either the single strandedness of the cleaved segment or the integrity of stem-loop I (Fig. 1b). We experimentally verified the structure of these RNAs using RNase T₁ (Fig. 4a), which specifically cleaves 3' to G residues in single-stranded regions¹⁵. RNase T₁

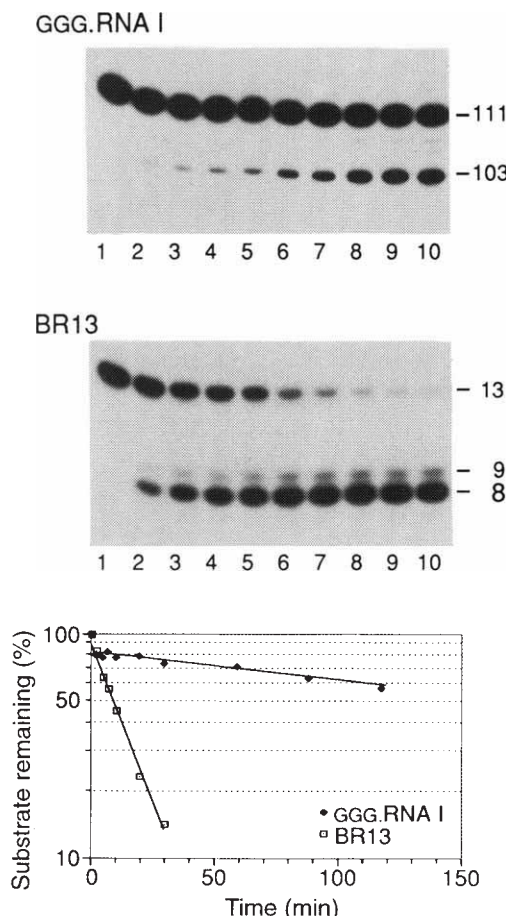


FIG. 3 Relative efficiency of RNase E cleavage of GGG.RNA I and oligoribonucleotide BR13. Continuously labelled GGG.RNA I (5 pmol; 4.6×10^{-3} A_{260} units) and 5 pmol (5.5×10^{-4} A_{260} units) of 5'-labelled oligoribonucleotide BR13 were incubated separately with 15 ng affinity-purified RNase E. Aliquots (4 μ l) were removed from each reaction (initial volume 50 μ l) after 0, 2, 5, 7, 10, 20, 30, 60, 90 and 120 min (lanes 1–10, respectively), mixed separately with 8 μ l sequencing stop buffer, denatured by incubating at 85 °C for 3 min, and run on 8% and 20% polyacrylamide sequencing gels, respectively. The semilogarithmic plot shows the amount of each substrate remaining as a function of time. From the gradient of each plot (supplementary information), we calculated that under these experimental conditions the time for RNase E to cleave 50% of oligoribonucleotide BR13 and GGG.RNA I was 10.6 ± 0.4 and 207.5 ± 37.3 min, respectively. Additionally, from the gradients (supplementary information) we determined the initial rates of cleavage of oligoribonucleotide BR13 and GGG.RNA I and calculated that the specific activity of the affinity-purified RNase E preparation for these substrates was 21 mU and 1 mU, respectively, per mg protein (1 U catalyses cleavage of 1 μ mol of substrate in 1 min). The same differential rates of decay were also obtained when the substrates were incubated with affinity-purified RNase E or a previously described RNase E preparation¹² under the conditions described in Fig. 4b.

METHODS. The conditions for RNase E cleavage and the procedures for labelling and purifying oligoribonucleotide BR13 were as described in Fig. 2. Continuously labelled GGG.RNA I was synthesized using T7 RNA polymerase from *HaeIII*-linearized pM21¹⁴ and purified from an 8% polyacrylamide sequencing gel as described previously¹³.

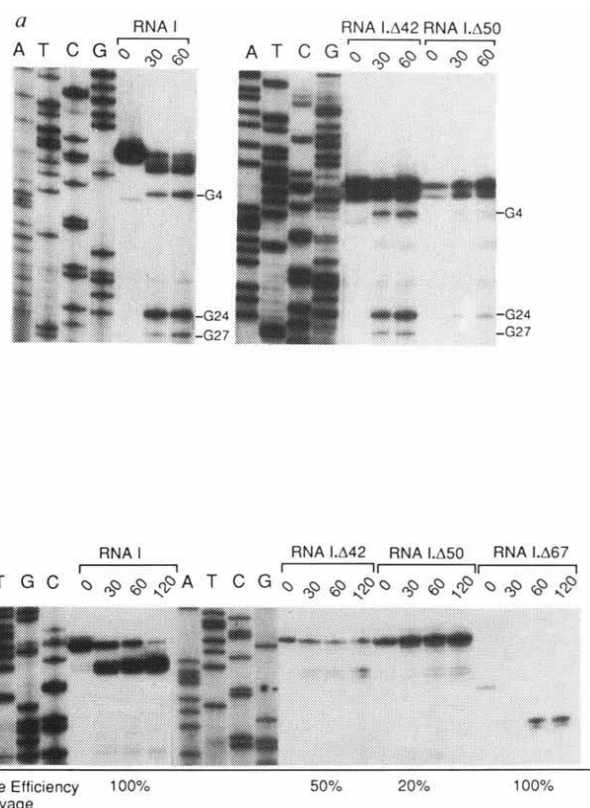


FIG. 4 RNase T₁ and RNase E cleavage of RNA I, RNA I.Δ42, RNA I.Δ50 and RNA I.Δ67. **a**, RNase T₁ analysis of RNA I variants containing tetranucleotide deletions in stem-loop II (Fig. 1). Horizontal lines indicate the positions of RNase T₁ cleavage mapped by primer extension²¹. RNA I variants were synthesized *in vitro* by T7 RNA polymerase; although, the transcripts consequently contained three additional G residues at their 5' end¹⁴, the nucleotides are numbered from the first A. RNA I.Δ67 was cleaved by RNase T₁ with the same efficiency, and at the same sites, as pBR322 RNA I (data not shown). **b**, Relative efficiency of RNase E cleavage of RNA I, RNA I.Δ42, RNA I.Δ50 and RNA I.Δ67. For each RNA I variant, the fraction cleaved by partially purified RNase E¹² after 30 min was determined and expressed relative to cleavage of wild-type RNA I under the same conditions. The gel shown contains unequal loadings to allow both the substrates and the corresponding cleavage products to be seen in a single exposure; however, as described previously¹², cleavage kinetics were determined from multiple exposures of a gel containing equal loadings of RNA. Substrates were total RNA samples isolated from exponentially growing cultures of the *rne*^{ts} *E. coli* strain N3431²² containing plasmids expressing RNA I, RNA I.Δ42, RNA I.Δ50 or RNA I.Δ67 30 min after shifting to the non-permissive temperature¹². **METHODS.** DNA segments encoding RNA I.Δ42, RNA I.Δ50, or RNA I.Δ67 were constructed by site-directed mutagenesis¹² of M13mp19ori²³ using oligodeoxynucleotides 5'-GCTACCAACTCTTTAAG-GTAACTGGCTTC, 5'-ATCAAGAGCTACCATTTTCCGAAGGTAAC, or 5'-GG-TGGTTTGTTCGTCAGAGCTACCAAC, respectively. Each deletion was confirmed by DNA sequencing and introduced into pCM128²⁴ as described previously¹² for the construction of the RNA I-expression plasmid pCML108. Templates for the *in vitro* synthesis of wild-type RNA I, RNA I.Δ42, RNA I.Δ50 and RNA I.Δ67 by T7 RNA polymerase were generated by PCR amplification of segments of the corresponding pCM128-derived plasmids using oligodeoxynucleotide primers, 5'-GGTACCTAATACGACTCACTATAGGACAGTATTGGTATCTGCGC (the sequence encoding a T7 promoter is in bold type) and 5'-ACAAAAAAC-CACCGCTACC. RNase T₁ cleavage reactions were done as described previously¹⁴ and the reaction products analysed by primer extension²¹. RNase E cleavage assay: substrates were incubated with partially purified RNase E as described previously¹². Oligonucleotide Su-29¹² was used to map sites of endonucleolytic cleavage in RNA I, RNA I.Δ42 and RNA I.Δ50 by primer extension (supplementary information); however, as the region deleted in RNA I.Δ67 overlaps the binding site for Su-29, cleavage sites in RNA I.Δ67 were mapped using the same oligodeoxynucleotide used to generate the DNA segment encoding this RNA I variant.

preferentially cleaved all of these mutant RNAs at positions G4, G24 and G27 (Fig. 4a), which are the same sites cleaved in native RNA I and in the GGG.RNA I variant¹⁴. However, the relative intensity of bands representing cleavage products was greatly reduced for RNA I.Δ50, suggesting that this RNA exists in a conformation that shields the single-stranded 5' segment and the loop of stem-loop I from cleavage by RNase T₁. Similarly, RNase E cleavage, which occurred in all of these RNA variants at the site cleaved in wild-type RNA I, was sharply decreased for RNA I.Δ50 (Fig. 4b). RNase E cleavage of RNA I.Δ42 was also reduced, although not to the same extent as cleavage of RNA I.Δ50 (Fig. 4b). These results suggest that overall substrate conformation influences the accessibility of RNase E, and other ribonucleases to potentially cleavable sites. However, in contrast to the effects we observed for RNase E and RNase T₁, the rate of cleavage by RNase V₁, which was used to confirm the integrity of the stem-loop adjacent to the RNase E cleavage site, was not affected by the mutations we studied (data not shown).

Our findings indicate that stem-loop regions in RNA I impede, rather than promote, RNase E cleavages, and that other modifications that affect higher-order structure can further reduce cleavage at susceptible sites. As RNase E does not require a stem-loop for entry to a substrate (this study) or a particular order of nucleotides^{12,13}, many (A+U)-rich sequences within complex RNAs are likely to contain sequences that can be cleaved, provided that the sequences are unpaired and are otherwise accessible to the enzyme. Stem-loop regions or alterations in overall substrate conformation may interfere with RNase E cleavages by limiting access of the enzyme. Additionally, our discovery that RNase E can attack short RNA oligomers raises the possibility that this endonuclease, which initiates the processing of ribosomal RNA precursors and the decay of a variety of other *E. coli* transcripts²⁻⁵, may also cut (at sites made available by the removal of regions of secondary structure) some of the fragments that it and other ribonucleases have generated by previous cleavages. Thus, RNase E cleavage resulting in or following stem-loop removal may occur reiteratively during the decay and processing of complex transcripts *in vivo*. If this is correct, the prolonged chemical half-life observed for bulk RNA and individual mRNA transcripts when RNase E activity is mutationally limited^{16,17} may reflect a decrease in endonucleolytic cleavage of decay intermediates as well as a reduced rate of cleavage of primary transcripts. □

Received 9 December 1994; accepted 24 January 1995.

- Misra, T. K. & Apirion, D. *J. biol. Chem.* **253**, 5594-5599 (1978).
- Ehretsmann, C. P., Carpousis, A. J. & Krusch, H. M. *FASEB J.* **6**, 3186-3192 (1992).
- Meleforts, O., Lundberg, U. & von Gabain, A. in *Control of Messenger RNA Stability* (eds Belasco, J. & Brawerman, G.) 53-70 (Academic, San Diego, 1993).
- Higgins, C. F., Peltz, S. W. & Jacobson, A. *Curr. Opin. Genet. Dev.* **2**, 739-747 (1992).
- Apirion, D., Miczak, A. & Taraseviciene, L. *Molec. Biol. (Life Sci. Adv.)* **11**, 221-233 (1992).
- Mackie, G. A. *J. biol. Chem.* **267**, 1054-1061 (1992).
- Bouvet, P. & Belasco, J. G. *Nature* **360**, 488-491 (1992).
- Cormack, R. S. & Mackie, G. A. *J. molec. Biol.* **228**, 1078-1090 (1992).
- Ehretsmann, C. P., Carpousis, A. J. & Krusch, H. M. *Genes Dev.* **6**, 149-159 (1992).
- Mackie, G. A. & Genereaux, J. L. *J. molec. Biol.* **234**, 998-1012 (1993).
- Tomcsanyi, T. & Apirion, D. *J. molec. Biol.* **185**, 713-720 (1985).
- Lin-Chao, S., Wong, T. T., McDowall, K. J. & Cohen, S. N. *J. biol. Chem.* **269**, 10797-10803 (1994).
- McDowall, K. J., Lin-Chao, S. & Cohen, S. N. *J. biol. Chem.* **269**, 10790-10796 (1994).
- Helmer-Citterich, M., Anceschi, M. M., Banner, D. W. & Cesareni, G. *EMBO J.* **7**, 557-566 (1988).
- Takahashi, K. *J. Biochem.* **49**, 1-8 (1961).
- Apirion, D. & Lassar, A. B. *J. biol. Chem.* **253**, 1738-1742 (1978).
- Ono, M. & Kuwano, M. *J. molec. Biol.* **129**, 343-357 (1979).
- Tamm, J. & Polisky, B. *Nucleic Acids Res.* **11**, 6381-6397 (1983).
- Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387-395 (1984).
- Hochuli, E., Dobeli, H. & Schacher, A. *J. Chromat.* **411**, 177-184 (1987).
- Lin-Chao, S., Chen, W. T. & Wong, T. T. *Molec. Microbiol.* **6**, 3385-3393 (1992).
- Goldsblum, K. & Apirion, D. *Cell* **15**, 1055-1066 (1981).
- Lin-Chao, S. & Cohen, S. N. *Cell* **65**, 1233-1242 (1991).
- Tucker, W. T., Miller, C. A. & Cohen, S. N. *Cell* **38**, 191-201 (1984).

ACKNOWLEDGEMENTS. This research was supported by grants from the National Science Council (N.S.C.) of the Republic of China to S. L.-C. and S.N.C. and from the U.S. National Institutes of Health to S.N.C. K.J.M. and V.R.K. received postdoctoral fellowships from the Science and Engineering Research Council of the United Kingdom and the N.S.C., respectively.

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London office of *Nature*.

ERRATUM

Interaction with RAP74 subunit of TFIIF is required for transcriptional activation by serum response factor

Véronique Joliot, Mark Demma & Ron Prywes

Nature **373**, 632-635 (1995).

In the contents page for the 16 February issue the title for this Letter included an error (TFIID for TFIIF) which could cause confusion. The Letter itself was printed correctly.

CORRECTIONS

γ-Interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes

Maria Gaczynska, Kenneth L. Rock & Alfred L. Goldberg

Nature **365**, 264-267 (1993)

THE caption to Fig. 1, page 265, lines 13-15, should have stated "one unit represents one μmol substrate cleaved per mg protein per hour" (not one nmol), and on lines 23-24, "*K_m* values of 0.4 mM and 0.5 mM" (not nM). Unfortunately, these typographical errors have led Ustrell *et al.* to argue that our findings were invalid, based on their incorrect impression that our kinetic values for proteasome varied widely between experiments and differed sharply from values obtained by others. However, calculations of the *V_{max}* and *K_m* from Figs 1 and 2 using the correct units yield values that resemble those in our other experiments (Tables 1-3) and fall within the range of values obtained by Ustrell *et al.*¹, by others²⁻⁴, and in our own subsequent work⁵. Moreover, the central finding in our paper (that γ-IFN alters the peptidase properties of the proteasome through the incorporation of the MHC-encoded subunits) has been confirmed in six different cell lines by three independent laboratories⁶⁻⁸, and also by our subsequent studies involving the transfection⁵ or deletion of the LMP genes⁹. □

- Ustrell, V., Pratt, G. & Rechsteiner, M. *Proc. natn. Acad. Sci. U.S.A.* **92**, 584-588 (1995).
- Ma, C. P., Slaughter, C. A. & DeMartino, G. N. *J. biol. Chem.* **267**, 10515-10523 (1992).
- Rivett, A. J. *J. biol. Chem.* **264**, 12215-12219 (1989).
- Boes, B. *et al. J. exp. Med.* **179**, 901-909 (1994).
- Gaczynska, M., Rock, K. L., Spies, T. & Goldberg, A. L. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9213-9217 (1994).
- Gaczynska, M., Rock, K. L. & Goldberg, A. L. *Nature* **365**, 264-267 (1993).
- Driscoll, J., Brown, M. G., Finley, D. & Monaco, J. J. *Nature* **365**, 262-264 (1993).
- Aki, M. *et al. J. Biochem.* **115**, 257-269 (1994).
- Van Kaer, L. *et al. Immunity* **1**, 533-541 (1994).

Design and synthesis of an ultraviolet-transparent nonlinear optical crystal Sr₂Be₂B₂O₇

Chuangtian Chen, Yebin Wang, Baichang Wu, Keche Wu, Wenlun Zeng & Linhua Yu

Nature **373**, 322-324 (1995)

THE name of the fourth author of this Letter was misspelt and is Kechen Wu. □